

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: July 19, 2001, 16:07:51 ; Search time 36.24 Seconds
(without alignments)
4472.231 Million cell updates/sec

Title: US-09-501-171A-4

Perfect score: 1225

Sequence: 1 MFARKPGCAPLGAMPVDPQ.....PYSELNYETSHYSPSPDSW 1225

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 425026 seqs, 132305027 residues

Word size: 0

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

SPREMBL.16:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_unclassified:*
13: sp_vertebrate:*
14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1225	100.0	1225	4 Q9UB3	Q9UB3 homo sapien
2	868	70.9	1040	4 Q9UM6	Q9UM6 homo sapien
3	775	63.3	876	4 Q4384	Q4384 homo sapien
4	720	58.8	1225	4 Q00379	Q00379 homo sapien
5	312	25.5	321	4 Q43206	Q43206 homo sapien
6	284	23.2	1247	11 Q35927	Q35927 mus musculu
7	264	21.6	264	11 Q35116	Q35116 rattus norv
8	206	16.8	307	4 Q15390	Q15390 homo sapien
9	114	9.3	114	4 Q9UPM3	Q9UPM3 homo sapien
10	26	2.1	1211	4 Q99569	Q99569 homo sapien
11	16	1.3	779	5 Q9W579	Q9W579 drosophila
12	16	1.3	781	5 Q9NHP1	Q9NHP1 drosophila
13	15	1.2	91	4 Q95645	Q95645 homo sapien
14	11	0.9	616	4 Q9UP73	Q9UP73 homo sapien
15	11	0.9	742	13 Q9DE61	Q9DE61 xenopus lae
16	11	0.9	838	4 Q9UP71	Q9UP71 homo sapien
17	11	0.9	907	13 Q9DFB2	Q9DFB2 xenopus lae
18	11	0.9	933	4 Q60935	Q60935 homo sapien
19	11	0.9	933	4 Q60935	Q60935 homo sapien

ALIGNMENTS

20	11	0.9	939	4	Q60714	Q60714 homo sapien
21	11	0.9	939	4	Q15088	Q15088 homo sapien
22	11	0.9	941	4	Q60713	Q60713 homo sapien
23	11	0.9	962	4	Q60715	Q60715 homo sapien
24	11	0.9	968	4	Q60716	Q60716 homo sapien
25	11	0.9	1181	5	Q9U308	Q9U308 caenorhabdi
26	10	0.8	607	10	Q9U308	Q9U308 caenorhabdi
27	10	0.8	1336	10	Q81074	Q81074 arabidopsis
28	9	0.7	53	4	Q92728	Q92728 homo sapien
29	9	0.7	83	5	Q9GT19	Q9GT19 heterodera
30	9	0.7	141	14	Q9DUC8	Q9DUC8 lt virus. O
31	9	0.7	153	5	Q9NA70	Q9NA70 caenorhabdi
32	9	0.7	239	2	Q66118	Q66118 zymomonas m
33	9	0.7	301	10	Q9LX93	Q9LX93 arabidopsis
34	9	0.7	368	5	Q45198	Q45198 caenorhabdi
35	9	0.7	400	10	Q9ZRM3	Q9ZRM3 brucea su
36	9	0.7	409	5	Q9ZRM3	Q9ZRM3 antilrhinum
37	9	0.7	415	6	Q9U027	Q9U027 leishmania
38	9	0.7	415	6	Q29015	Q29015 sus scrofa
39	9	0.7	443	5	Q9XVA8	Q9XVA8 dictyosteli
40	9	0.7	446	4	Q15881	Q15881 homo sapien
41	9	0.7	451	11	Q62775	Q62775 rattus norv
42	9	0.7	477	6	Q97600	Q97600 oryctolagus
43	9	0.7	479	11	Q9JMK4	Q9JMK4 mus musculu
44	9	0.7	485	11	Q920C8	Q920C8 rattus norv
45	9	0.7	496	4	Q9UDY7	Q9UDY7 homo sapien

RESULT 1
Q9UB3 PRELIMINARY: PRT: 1225 AA.

AC Q9UB3;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE NEURAL PLAKOPHILIN-RELATED ARM-REPEAT PROTEIN (NPRAP).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE=99223289; PubMed=10208590;
RA Tanahashi H.;
RT "Isolation of human delta-catenin and its binding specificity with
RL presentin 1.";
RL NeuroReport 10:563-568(1999).
DR EMBL: AB013805; BAA36167-17;
DR InterPro: IPR000225;
DR Pfam: PF00514; Armadillo_seg. 6.
DR PROSITE: PS00176; ARM_REPEAT. 3.
DR SMART: SM00185; ARM. 1
SQ SEQUENCE 1225 AA; 13265 MW; 4A24AC3E22BFE83 CRC64;

2-25-1999
post positivity check

Query Match 100.0%; Score 1225; DB 4; Length 1225;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MFARKPGCAPLGAMPVDPQSSASEKTSLSFGLTSGDSEETTTAIIASVKEDEL 60
Db 1 MFARKPGCAPLGAMPVDPQSSASEKTSLSFGLTSGDSEETTTAIIASVKEDEL 60
QY 61 QFERLRLFAERQIVASQERCKLSFGSGMSMSASAEQOMODOKIDIELDTTG 120
Db 61 QFERLRLFAERQIVASQERCKLSFGSGMSMSASAEQOMODOKIDIELDTTG 120
QY 121 LEIVDSCTIRLSQGLIDPDYGTGERPSILSOSALQNSKEGFSFOYASVNSQTAL 180
Db 121 LEIVDSCTIRLSQGLIDPDYGTGERPSILSOSALQNSKEGFSFOYASVNSQTAL 180

Db 121 LELVDSICIRLQESGILUPDQYSTGERPSILSQSALQJLNSKPEGSFOYPASYSNQTAL 180
 QY 181 GETTSPOLPARQTOARATGOSOSTTSRAGHLAGPEPAPPPPPPPPPPPPPPPPPPPPPPP 240
 Db 181 GETTSPOLPARQTOARATGOSOSTTSRAGHLAGPEPAPPPPPPPPPPPPPPPPPPPPPPP 240
 QY 241 PDAPPAALAAALYSSSTLPAPPRGSPPLAOPGSPKTLORGSGAPGATYAAPRGSSP 300
 Db 241 PDAPPAALAAALYSSSTLPAPPRGSPPLAOPGSPKTLORGSGAPGATYAAPRGSSP 300
 QY 301 KOSPERLAKSTSTSPINIVYSSAGLSPIRYTSPVOSTISSPIHQLSSTICTYATLS 360
 Db 301 KOSPERLAKSTSTSPINIVYSSAGLSPIRYTSPVOSTISSPIHQLSSTICTYATLS 360
 QY 361 PTKRLVHASSEYOSKSHSOLYATATLORPGSLAAGSRASYSQHGHLGELRALOSPEHHT 420
 Db 361 PTKRLVHASSEYOSKSHSOLYATATLORPGSLAAGSRASYSQHGHLGELRALOSPEHHT 420
 QY 421 DEYEDRYOKRPMKSLISQSGDPLPAHNTGTYRTSTAPSSPVGYDVPLORTSGOHGPON 480
 Db 421 DEYEDRYOKRPMKSLISQSGDPLPAHNTGTYRTSTAPSSPVGYDVPLORTSGOHGPON 480
 QY 481 AAAATFOASVYAGPASYADPYRQLOCYCPESVPSYKSGPALPEEGTLARSPSIDSIQK 540
 Db 481 AAAATFOASVYAGPASYADPYRQLOCYCPESVPSYKSGPALPEEGTLARSPSIDSIQK 540
 QY 541 DPREGMDPELPEYIOMLOHQFPVSQNAAYLOHLCFGDNKTKAETIRROGGIQLLYDL 600
 Db 541 DPREGMDPELPEYIOMLOHQFPVSQNAAYLOHLCFGDNKTKAETIRROGGIQLLYDL 600
 QY 601 LDHMTVEVHSACGALNLYYKANDDNKIALKNGGIPALVRLRKTDTLEIRELVTVG 660
 Db 601 LDHMTVEVHSACGALNLYYKANDDNKIALKNGGIPALVRLRKTDTLEIRELVTVG 660
 QY 661 LWNSSCALMPILIOALAVLTNAVLIIPHSGWNSPLQDRKTOHSSQVLRNATGCLR 720
 Db 661 LWNSSCALMPILIOALAVLTNAVLIIPHSGWNSPLQDRKTOHSSQVLRNATGCLR 720
 QY 721 NVSSAGEBARRECDGLDALLVYIOSALGSSEIDSKTEVENCVCILRNLSTRLAETS 780
 Db 721 NVSSAGEBARRECDGLDALLVYIOSALGSSEIDSKTEVENCVCILRNLSTRLAETS 780
 QY 781 OGHMGDELGLGCGANGDASSCGCKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 840
 Db 781 OGHMGDELGLGCGANGDASSCGCKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 840
 QY 841 LHMPSIVKPYLTLLSECSNPDTEGAAGALQNLAAAGSMKMSVYIRAVRREKGPILVEL 900
 Db 841 LHMPSIVKPYLTLLSECSNPDTEGAAGALQNLAAAGSMKMSVYIRAVRREKGPILVEL 900
 QY 901 LRINDRYVCAVATLRNMLDVANKELIGYAMRDLYHRLPGGNNNSNTASKAMSDTY 960
 Db 901 LRINDRYVCAVATLRNMLDVANKELIGYAMRDLYHRLPGGNNNSNTASKAMSDTY 960
 QY 961 TAVCCTIHEVITKMEKAKALRDAGIEKLVGSKSGDKHSPKYVAAQVLTNSMWOYR 1020
 Db 961 TAVCCTIHEVITKMEKAKALRDAGIEKLVGSKSGDKHSPKYVAAQVLTNSMWOYR 1020
 QY 1021 DLRSVYKDDMSOYHFAVASSSTIERORPYSSSRTPSISPVAVSPNNRSASAPSPREM 1080
 Db 1021 DLRSVYKDDMSOYHFAVASSSTIERORPYSSSRTPSISPVAVSPNNRSASAPSPREM 1080
 QY 1081 ISLKERKTEVETGCSNATYGAKEGHTSRKDAVTAONTGISTLYRNSYGAPADIKHNOY 1140
 Db 1081 ISLKERKTEVETGCSNATYGAKEGHTSRKDAVTAONTGISTLYRNSYGAPADIKHNOY 1140
 QY 1141 SAOPVPEPSRKDYETQOPONSTRNVDSEFFEDOVHHRPASEYMHGIGKSTGVYDF 1200
 Db 1141 SAOPVPEPSRKDYETQOPONSTRNVDSEFFEDOVHHRPASEYMHGIGKSTGVYDF 1200
 QY 1201 YSAARPYSELNYSHTSHYPASDSMV 1225
 Db 1201 YSAARPYSELNYSHTSHYPASDSMV 1225

RESULT 2
 Q90M66 ID Q90M66 PRELIMINARY; PRT: 1040 AA.
 AC Q90M66
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE G124 (FRAGMENT).
 GN G124.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_TaxId=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Levesque G., Yu G., Fraser P.E., StGeorge-Hyslop P.;
 RT "Presentin 1 interacts with a novel protein that contains armadillo
 RT repeats and maps near the Cri du chat locus on chromosome 5p.";
 RL Submitted (DEC-1996) to the EMBL/Genbank/DBJ databases.
 DR EMBL: U61004; AAD00453.1; -.
 DR InterPro: IPR000225; -.
 DR Pfam: PF00514; Armadillo_seg; 6.
 DR PROSITE: PS0176; ARM_REPEAT; 3.
 DR SMART: SM00185; ARM; 1.
 FT NON_TER 1
 SQ SEQUENCE 1040 AA; 112802 MW; E12C8694EAC51D42 CRC64;

Query Match 70.9%; Score 868; DB 4; Length 1040;
 Best local Similarity 99.9%; Pred. No. 0;
 Matches 968; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 257 STLPAPRGSPPLAOPGSPKTLORGSGAPGATYAAPRGSSPKQSPRLAKSTSTSP 316
 Db 72 STLPAPRGSPPLAOPGSPKTLORGSGAPGATYAAPRGSSPKQSPRLAKSTSTSP 131
 QY 317 INIVSSAGLSPIRYTSPVOSTISSPIHQLSSTICTYATLSPTKRLVHASSEYOSKHS 376
 Db 132 INIVSSAGLSPIRYTSPVOSTISSPIHQLSSTICTYATLSPTKRLVHASSEYOSKHS 191
 QY 377 QELYATATLORPGSLAAGSRASYSQHGHLGELRALOSPEHHTDEYEDRYOKRPMKSL 436
 Db 192 QELYATATLORPGSLAAGSRASYSQHGHLGELRALOSPEHHTDEYEDRYOKRPMKSL 251
 QY 437 LSQSGDPLPAHNTGTYRTSTAPSSPVGYDVPLORTSGOHGPQNAATFOASVYAGPA 496
 Db 252 LSQSGDPLPAHNTGTYRTSTAPSSPVGYDVPLORTSGOHGPQNAATFOASVYAGPA 311
 QY 497 SNYADPYRQLOCYCPESVPSYKSGPALPEEGTLARSPSIDSIQKDPREFGMDPELPEVI 556
 Db 312 SNYADPYRQLOCYCPESVPSYKSGPALPEEGTLARSPSIDSIQKDPREFGMDPELPEVI 371
 QY 557 OMLQHOFPVSQNAAYLOHLCFGDNKTKAETIRROGGIQLLYDLIDHMTVEVHSACGAL 616
 Db 372 OMLQHOFPVSQNAAYLOHLCFGDNKTKAETIRROGGIQLLYDLIDHMTVEVHSACGAL 431
 QY 617 RVLVYKANDDNKIALKNGGIPALVRLRKTDTLEIRELVTVGLWLNSSCDALKMPITIQ 676
 Db 432 RVLVYKANDDNKIALKNGGIPALVRLRKTDTLEIRELVTVGLWLNSSCDALKMPITIQ 491
 QY 677 DALAVLTNAVLIIPHSGWNSPLQDRKTOHSSQVLRNATGCLRNVSSAGEBARREKREC 736
 Db 492 DALAVLTNAVLIIPHSGWNSPLQDRKTOHSSQVLRNATGCLRNVSSAGEBARREKREC 551
 QY 737 DGLTDALLVYIOSALGSSEIDSKTEVENCVCILRNLSTRLAETSQGHMGTDDELGLGCG 796
 Db 552 DGLTDALLVYIOSALGSSEIDSKTEVENCVCILRNLSTRLAETSQGHMGTDDELGLGCG 611
 QY 797 EANGKDASSCGCKKK 856
 Db 612 EANGKDASSCGCKKK 671

```

OY 857 CSNPDTLEGAAGALONTAAGSMKMSVYIRAAVRKEKGLPIVELLRIDNDRVCAVATATL 916
DB 672 CSNPDTLEGAAGALONTAAGSMKMSVYIRAAVRKEKGLPIVELLRIDNDRVCAVATATL 731
OY 917 RNMAADVRRKELIGKYMADLVHRLPGNNSNNTASKANSDDTVAACCTLHEVITTKNE 976
DB 732 RNMAADVRRKELIGKYMADLVHRLPGNNSNNTASKANSDDTVAACCTLHEVITTKNE 791
OY 977 NAKALDAGIEKLVGISKDGKHSKPKVYKAAQVLTNSMOMQYRDLRSYKXKDGMSOYHF 1036
DB 792 NAKALDAGIEKLVGISKDGKHSKPKVYKAAQVLTNSMOMQYRDLRSYKXKDGMSOYHF 851
OY 1037 VASSTTERDROPRYSSTSPISPRVSPNNRSASAPREMIISLKRKTDYECTGSN 1096
DB 852 VASSTTERDROPRYSSTSPISPRVSPNNRSASAPREMIISLKRKTDYECTGSN 911
OY 1097 ATYHGAKEHTSRKDMTAONTGISTLYNSYGAPAEADIKHNOVSAQVPQEPSRKDYET 1156
DB 912 ATYHGAKEHTSRKDMTAONTGISTLYNSYGAPAEADIKHNOVSAQVPQEPSRKDYET 971
OY 957 YOPFONSTRNYDSFFEDQVHHRPPASEYTMHLGLKSTGNYDVFYSARPSELNYESH 1216
DB 972 YOPFONSTRNYDSFFEDQVHHRPPASEYTMHLGLKSTGNYDVFYSARPSELNYESH 1031
OY 1217 YPASPSDW 1225
DB 1032 YPASPSDW 1040

RESULT 3
OY 043840 PRELIMINARY: PRT: 876 AA.
AC 043840;
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE NEURAL PLAKOPHTILIN RELATED ARM-REPEAT PROTEIN (FRAGMENT).
GN NRPAP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98002299; PubMed=9342840;
RA Paffenholz R., Franke W.W.;
RT "Identification and localization of a neurally expressed member of the
RT plakoglobin/armadillo multigene family.";
RL Differentiation 61:293-304(1997).
DB EMBL: U52351; AAB97957.1; -
DR InterPro: IPR000225; -
DR Pfam: PF00514; Armadillo_seg. 6.
DR PROSITE: PSS0176; ARM_REPEAT. 3.
DR SMART: SM00185; ARM. 1.
FT NON_TER 1
FT SEQUENCE 876 AA; 96442 MW; 8D2342B71CC9E92A CRC64;
SQ

```

```

DB 121 QRTSOHQHPQNAATAATFORASVAGPASNVADPYROLQCPVSPYSKSGPALPREGTL 180
OY 530 ARSPSIDSIQDPRFEGWRDPELPEVIOMLQHPFVSQVNAAYLOHLCGDNKIKAEIR 589
DB 181 ARSPSIDSIQDPRFEGWRDPELPEVIOMLQHPFVSQVNAAYLOHLCGDNKIKAEIR 240
OY 590 ROGGIQLLVLDLDRHMTVEHRSACGALRNLYGKANDDNKIALKNCGGITAYRLKRT 649
DB 241 ROGGIQLLVLDLDRHMTVEHRSACGALRNLYGKANDDNKIALKNCGGITAYRLKRT 300
OY 650 DLEIRELVTVGLNMLSSCDALKNPIODALAVLTNAVITPHSGWENSPLODDRKIQHSS 709
DB 301 DLEIRELVTVGLNMLSSCDALKNPIODALAVLTNAVITPHSGWENSPLODDRKIQHSS 360
OY 710 QVLRNATGCLNNSASGEARRRRECGLTDALLYIOSALGSEISDSKTVENCYILR 769
DB 361 QVLRNATGCLNNSASGEARRRRECGLTDALLYIOSALGSEISDSKTVENCYILR 420
OY 770 NLSTYRLAETSOQOHMGTDDELGLGCBANGKDAESSGCMGKKKKKSDQMDGVGLP 829
DB 421 NLSTYRLAETSOQOHMGTDDELGLGCBANGKDAESSGCMGKKKKKSDQMDGVGLP 480
OY 830 DCAEPPKGIOMLMPSTYKPYTLTLLSECSNPDTLEGAAGALONTAAGSMKMSVYIRAAVR 889
DB 481 DCAEPPKGIOMLMPSTYKPYTLTLLSECSNPDTLEGAAGALONTAAGSMKMSVYIRAAVR 540
OY 890 KEGGLPIVELLRIDNDRVCAVATATLERNMALDYNKELIKYMRDLYHRLPGNNSNN 949
DB 541 KEGGLPIVELLRIDNDRVCAVATATLERNMALDYNKELIKYMRDLYHRLPGNNSNN 600
OY 950 TASKAMSDDTVAACCTLHEVITTKMENAKALRDAGIEKLVGISKDGKHSKPKVYKAA 1009
DB 601 TASKAMSDDTVAACCTLHEVITTKMENAKALRDAGIEKLVGISKDGKHSKPKVYKAA 660
OY 1010 SOVLNSMOMQYRDLRSYKXKDGMSOYHFVASSSTTERDROPRYSSTSPISPRVSPNNR 1069
DB 661 SOVLNSMOMQYRDLRSYKXKDGMSOYHFVASSSTTERDROPRYSSTSPISPRVSPNNR 720
OY 1070 SASAPASPREMIISLKRKTDYECTGSNATYHGAKEHTSRKDMTAONTGISTLYNSYG 1129
DB 721 SASAPASPREMIISLKRKTDYECTGSNATYHGAKEHTSRKDMTAONTGISTLYNSYG 780
OY 1130 APAEDIKHNOVSAQVPQEPSRKDYETQOPFONSTRNYDSFFEDQVHHRPPASEYTMHL 1189
DB 781 APAEDIKHNOVSAQVPQEPSRKDYETQOPFONSTRNYDSFFEDQVHHRPPASEYTMHL 840
OY 1190 GLKSTGNYDVFYSARPSELNYESHYPASPSDW 1225
DB 841 GLKSTGNYDVFYSARPSELNYESHYPASPSDW 876

RESULT 4
OY 000379 PRELIMINARY: PRT: 1225 AA.
AC 000379; Q13589;
DT 01-JUL-1997 (Tremblrel. 04, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE DELTA-CATENIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=97366296; PubMed=9223106;
RA Zhou J., Iiyangage U., Medina M., Ho C., Simmons A.D., Kosik K.S.;
RT "Presentin 1 interaction in the brain with a novel member of the
RT Armadillo family.";
RL NeuroReport 8:2085-2090(1997).
RN [2]
RP SEQUENCE FROM N.A.

```

RC TISSUE-BRAIN;
 RA Lu Q., Paredes M., Medina M., Zhou J., Cavallo R., Peifer M.,
 RA Orecchio L., Kosik K.S.;
 RT "d-carehin, an adhesive junction associated protein which promotes
 RT motile behavior";
 RL Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 865-1225 FROM N.A.
 RA MEDLINE=97202103; PubMed=9049630;
 RX Simmons A.D., Overhauser J., Loyett M.;
 RT "Isolation of cDNAs from the Cri-du-chat critical region by direct
 RT screening of a chromosome 5-specific cDNA library";
 RL Genome Res. 7:118-127(1997).
 DR EMBL; U96136; AAC63103.1; -;
 DR EMBL; U52828; AAB96357.1; -;
 DR InterPro: IPR000225; -;
 DR Pfam: PF00514; Armadillo, seg. 6.
 DR PROSITE: PS0176; ARM_REPEAT; 3.
 DR SMART: SM0185; ARM; 1.
 SQ SEQUENCE 1225 AA; 132664 MW; 8B676CFD5AFA4E65 CRC64;

Query Match 58.8%; Score 720; DB 4; Length 1225;

Best Local Similarity 99.6%; Pred. No. 0;
 Matches 1220; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MFARKPGAGAPGAPYDQSSASEKTSLSPLANTSGDSEETTSAILASVKEQL 60
 Db 1 MFARKPGAGAPGAPYDQSSASEKTSLSPLANTSGDSEETTSAILASVKEQL 60
 QY 61 QEEELTELEERIVASOLEKCLGSETSMSSMAEEDFOWOSODGKIDEDLTG 120
 Db 61 QEEELTELEERIVASOLEKCLGSETSMSSMAEEDFOWOSODGKIDEDLTG 120
 QY 121 LEVDSCIRLSQESGILDPDYSTERPILLSOSALQLNKPEGSFQYPASYHSNOTLAL 180
 Db 121 LEVDSCIRLSQESGILDPDYSTERPILLSOSALQLNKPEGSFQYPASYHSNOTLAL 180
 QY 181 GETTSOLPARGTQARATGGSFSGTTSRAGHLAAGPEAPPPPPPREPPAPSLGSAFHL 240
 Db 181 GETTSOLPARGTQARATGGSFSGTTSRAGHLAAGPEAPPPPPPREPPAPSLGSAFHL 240
 QY 241 PDAPPAALAAALYSSSTLPAPRGSSPLAAPGGSTPKLQREGSAPEGATTYAPGSSP 300
 Db 241 PDAPPAALAAALYSSSTLPAPRGSSPLAAPGGSTPKLQREGSAPEGATTYAPGSSP 300
 QY 301 KQSPSLAKSYSTSSPINIVSSAGSLPIRVTPPVQOSTISSPIHQLSSTIGTYATLS 360
 Db 301 KQSPSLAKSYSTSSPINIVSSAGSLPIRVTPPVQOSTISSPIHQLSSTIGTYATLS 360
 QY 361 PTKRLVHASQYKSHQELATATLQPGSLAAGSRAVSQGHGPELALQSPENHI 420
 Db 361 PTKRLVHASQYKSHQELATATLQPGSLAAGSRAVSQGHGPELALQSPENHI 420
 QY 421 DPYEDVYQKPMRSLSOSQDPLPAHNGYRTSTAPSSPEVDVPLQRTGSQGPON 480
 Db 421 DPYEDVYQKPMRSLSOSQDPLPAHNGYRTSTAPSSPEVDVPLQRTGSQGPON 480
 QY 481 AAATFORASAYAPASNVADPYROLQCPVSPYSKSGPALPREGTILARSPIDSIOK 540
 Db 481 AAATFORASAYAPASNVADPYROLQCPVSPYSKSGPALPREGTILARSPIDSIOK 540
 QY 541 DPREFGWRDPELVYOMLOHOPPSQSNAAAYLQHLCEGDNKIKAEIRROGCIQLVLD 600
 Db 541 DPREFGWRDPELVYOMLOHOPPSQSNAAAYLQHLCEGDNKIKAEIRROGCIQLVLD 600
 QY 601 LDRHMEVHSAGALBNLYGKANDNKITALKNGGISTALVRLKTTDLDELRELVYGV 660
 Db 601 LDRHMEVHSAGALBNLYGKANDNKITALKNGGISTALVRLKTTDLDELRELVYGV 660
 QY 661 LWNLSGCDALKMPDIODALAVLITNAVLIPIHSGWENSPLODDKRIQLHSQVLRNATGCLR 720
 Db 661 LWNLSGCDALKMPDIODALAVLITNAVLIPIHSGWENSPLODDKRIQLHSQVLRNATGCLR 720

QY 721 NVSSAGEARARRRRECDGTLTDLALVYOSALSSSEIDSKTYENCYCILRNLSYRLAETS 780
 Db 721 NVSSAGEARARRRRECDGTLTDLALVYOSALSSSEIDSKTYENCYCILRNLSYRLAETS 780
 QY 781 QGQHMGTDELGLLGEANGKDAESSGCGKKKKKKKSODQMDGVPDPDCAEPKGIOM 840
 Db 781 QGQHMGTDELGLLGEANGKDAESSGCGKKKKKKKSODQMDGVPDPDCAEPKGIOM 840
 QY 841 LHPSTIVKPYLLTLLSECSNPDTLEGAAGALQNLAAAGSMKWSYIIRAAVRKEGPIPIVEL 900
 Db 841 LHPSTIVKPYLLTLLSECSNPDTLEGAAGALQNLAAAGSMKWSYIIRAAVRKEGPIPIVEL 900
 QY 901 LRIIDNRVCAVATLRNALDVRKELIGKXAMDLYHRLPGGNSNNTASKAMSDTV 960
 Db 901 LRIIDNRVCAVATLRNALDVRKELIGKXAMDLYHRLPGGNSNNTASKAMSDTV 960
 QY 961 TAVCCTLHEVITKMMENKALRDAGIEKLVGISKSKDKHSPPKVKAAQVLSMMQYR 1020
 Db 961 TAVCCTLHEVITKMMENKALRDAGIEKLVGISKSKDKHSPPKVKAAQVLSMMQYR 1020
 QY 1021 DLRSLYKKGMSQYHFVASSSTIERDQRPYSSSTPSISPVRSVNNRSASAPASPREM 1080
 Db 1021 DLRSLYKKGMSQYHFVASSSTIERDQRPYSSSTPSISPVRSVNNRSASAPASPREM 1080
 QY 1081 ISLKERKTDYECTGSNATYHGAKGHTSRKDMTQONTGISITVRSYGAPAEIKHNOV 1140
 Db 1081 ISLKERKTDYECTGSNATYHGAKGHTSRKDMTQONTGISITVRSYGAPAEIKHNOV 1140
 QY 1141 SAQPVQPEPSRKDYETYPQFONSTRNDESEFEDVHHRPASEYTMHLGLKSTGVYDF 1200
 Db 1141 SAQPVQPEPSRKDYETYPQFONSTRNDESEFEDVHHRPASEYTMHLGLKSTGVYDF 1200
 QY 1201 YSARPYSELNETSHYPASPSW 1225
 Db 1201 YSARPYSELNETSHYPASPSW 1225

RESULT 5
 ID 043206 PRELIMINARY; PRT; 321 AA.
 AC 043206;
 DT 01-JUN-1998 (TRENDEL. 06, Created)
 DT 01-JUN-1998 (TRENDEL. 06, Last sequence update)
 DE 01-MAY-2000 (TRENDEL. 13, Last annotation update)
 DE PROBABLE CATEININ.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE=96207227; PubMed=8619474;
 RA Anderson B., Wentland M.A., Ricafrente J.Y., Liu W., Gibbs R.A.;
 RT "A 'double adaptor' method for improved shotgun library
 RT construction";
 RL Anal. Biochem. 236:107-113(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE=97264341; PubMed=9110174;
 RA Yu W., Anderson B., Worley K.C., Muzny D.M., Ding Y., Liu W.,
 RA Ricafrente J.Y., Wentland M.A., Lennon G., Gibbs R.A.;
 RT "Large-scale concatenation cDNA sequencing";
 RL Genome Res. 7:353-358(1997).
 DR EMBL; AF035302; AAB88185.1; -;
 KW Cytoskeleton; Structural protein; Cell adhesion.
 SQ SEQUENCE 321 AA; 36108 MW; 1DDF4811891DA053 CRC64;

Query Match 25.5%; Score 312; DB 4; Length 321;
 Best Local Similarity 100.0%; Pred. No. 0;

Matches 312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 914 TAIKRNALDVRNKLIGKAYMRDLVRLDGGNNNTASKAMDDVTVACCTLHEVTR 973
 DB 10 TAIKRNALDVRNKLIGKAYMRDLVRLDGGNNNTASKAMDDVTVACCTLHEVTR 69
 QY 974 NMENAKALRDAGIEKLVGSKSGDKHSPKVVKAASQVLYNSMWQYRDLRSYLKKKGWSQ 1033
 DB 70 NMENAKALRDAGIEKLVGSKSGDKHSPKVVKAASQVLYNSMWQYRDLRSYLKKKGWSQ 129
 QY 1034 YHFAVSSSTIERDORPYSSTSPISPVAVSPNNRSASAPASPREMISLKEKTDYECT 1093
 DB 130 YHFAVSSSTIERDORPYSSTSPISPVAVSPNNRSASAPASPREMISLKEKTDYECT 189
 QY 1094 GSNATYGAAGEHTRSRKADMTAONTGISTLRNSYGAPADIDKHNOVASOPVQEPSRKD 1153
 DB 190 GSNATYGAAGEHTRSRKADMTAONTGISTLRNSYGAPADIDKHNOVASOPVQEPSRKD 249
 QY 1154 YETVOPQNSTRNVDSEFEDQVHHRPASEYTMHLGKSTGVYVDFYSARPYSSELYE 1213
 DB 250 YETVOPQNSTRNVDSEFEDQVHHRPASEYTMHLGKSTGVYVDFYSARPYSSELYE 309
 QY 1214 TSHYPASPDMSV 1225
 DB 310 TSHYPASPDMSV 321

RESULT 6
 ID 035927 PRELIMINARY; PRT: 1247 AA.
 AC 035927;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE NEURAL PLAKOPHILIN-RELATED ARM-REPEAT PROTEIN.
 GN CTRND2 OR NEBPAP.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RX MEDLINE=98002299; PubMed=9342840;
 RA Pattenholz R., Franke W.W.;
 RT "Identification and localization of a neurally expressed member of the
 RL plakoglobin/armadillo multigene family.";
 RT Differentiation 61:293-304(1997).
 DR EMBL, U90331; AAB82409.1; -;
 DR SD, MG1:1195966; Ctrnd2.
 DR InterPro: IPR000225; -;
 DR Pfam: PF00514; Armadillo_seg; 6.
 DR PROSITE: PS50176; ARM_REPEAT; 3.
 DR SMART: SM00185; ARM; 1.
 SQ SEQUENCE 1247 AA; 134998 MW; DA7A6B6A27D2919 CRC64;

Query Match 23.2%; Score 284; DB 11; Length 1247;
 Best Local Similarity 100.0%; Pred. No. 2.3e-278;
 Matches 284; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 514 SPYKSGPALPPEGLTARSPSIDSIQKDPREFGWRDPLPEVITOMLQHOPSVOSNAAY 573
 DB 511 SPYKSGPALPPEGLTARSPSIDSIQKDPREFGWRDPLPEVITOMLQHOPSVOSNAAY 570
 QY 574 LOHLCFGNKKIKAEIRROGGIQLVLDLHDHMTVEHRSAGALRNLYGKANDNKITALK 633
 DB 571 LOHLCFGNKKIKAEIRROGGIQLVLDLHDHMTVEHRSAGALRNLYGKANDNKITALK 630
 QY 634 NCGGIPALVRLRKTDTLEIRELVTVGLVNLSSCDALKMPIIDALAVLNAVVIIPHSGW 693
 DB 631 NCGGIPALVRLRKTDTLEIRELVTVGLVNLSSCDALKMPIIDALAVLNAVVIIPHSGW 690

QY 694 ENSPLQDDRKIQLHSSQVLRNATGCLRNVSAGGEARRRMECDGLTDALLYVQSALGS 753
 DB 691 ENSPLQDDRKIQLHSSQVLRNATGCLRNVSAGGEARRRMECDGLTDALLYVQSALGS 750
 QY 754 SEIDSKTYENCVCILRNLSYRLAETSGOGHMGTDGLGCGE 797
 DB 751 SEIDSKTYENCVCILRNLSYRLAETSGOGHMGTDGLGCGE 794

RESULT 7
 ID 035116 PRELIMINARY; PRT: 264 AA.
 AC 035116;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE DELTA-CATENIN (FRAGMENT).
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxId=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MISTAR; TISSUE=BRAIN;
 RA Tananashi H.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL, AB008752; BAA23384.1; -;
 DR InterPro: IPR000225; -;
 DR Pfam: PF00514; Armadillo_seg; 2.
 DR PROSITE: PS50176; ARM_REPEAT; 2.
 DR SMART: SM00185; ARM; 1.
 FT NON_TER 1 264
 FT NON_TER 264 264
 SQ SEQUENCE 264 AA; 28928 MW; 065A86B8BD2F58E0 CRC64;

Query Match 21.6%; Score 264; DB 11; Length 264;
 Best Local Similarity 100.0%; Pred. No. 1.1e-258;
 Matches 264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 563 FPSVQSNAAAYLQHLFCFGDNKIKAEIRROGGIQLVLDLHDHMTVEHRSAGALRNLYG 622
 DB 1 FPSVQSNAAAYLQHLFCFGDNKIKAEIRROGGIQLVLDLHDHMTVEHRSAGALRNLYG 60
 QY 623 KANDDNKIKLKNCGGIPALVRLRKTDTLEIRELVTVGLVNLSSCDALKMPIIDALAVL 682
 DB 61 KANDDNKIKLKNCGGIPALVRLRKTDTLEIRELVTVGLVNLSSCDALKMPIIDALAVL 120
 QY 683 TNVVIIPHSGWENSPLODDRKIQLHSSQVLRNATGCLRNVSAGGEARRRMECDGLTDA 742
 DB 121 TNVVIIPHSGWENSPLODDRKIQLHSSQVLRNATGCLRNVSAGGEARRRMECDGLTDA 180
 QY 743 LLYVQSALGSSSEIDSKTYENCVCILRNLSYRLAETSGOGHMGTDGLGCGANGKD 802
 DB 181 LLYVQSALGSSSEIDSKTYENCVCILRNLSYRLAETSGOGHMGTDGLGCGANGKD 240
 QY 803 AESSGCWGKKKKKKKSQDQWDG 826
 DB 241 AESSGCWGKKKKKKKSQDQWDG 264

RESULT 8
 ID 015390 PRELIMINARY; PRT: 307 AA.
 AC 015390;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-AUG-1998 (TREMBLrel. 07, Last annotation update)
 DE GTR24.
 GN GTR24.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RA Frisner P.E., Levesque G., Rogueva E.A., Yu G., St George-Hyslop P.H.;
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U72655; AAB6859.1; -
 SQ SEQUENCE 307 AA; 34417 MW; 35640DA6113E65F CRC64;

Query Match 16.8%; Score 206; DB 4; Length 307;
 Best Local Similarity 99.7%; Pred. No. 7.8e-200;
 Matches 306; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 919 MALDVRNKLKIKYAMRDLVHRLPGNNNSNTASKAMSDDTVTAVCCCTLHEVITKNNENA 978
 DB 1 MALDVRNKLKIKYAMRDLVHRLPGNNNSNTASKAMSDDTVTAVCCCTLHEVITKNNENA 60
 QY 979 KALRDAGGIEKLVGISKSGDKHSPKVKAAQVILNSMWOYRDLRLSLYKKDQMSQYHFA 1038
 DB 61 KALRDAGGIEKLVGISKSGDKHSPKVKAAQVILNSMWOYRDLRLSLYKKDQMSQYHFA 120
 QY 1039 SSSTTEROROPYSSTPSPISPVAVSPNNRSASAPSPREMIISLKEKTDYECTGSNAT 1098
 DB 121 SSSTTEROROPYSSTPSPISPVAVSPNNRSASAPSPREMIISLKEKTDYECTGSNAT 180
 QY 1099 YHAGKEHTSRKDMTAQNTGISTLYRNSYGAPADIKHNOVSAQPPQEPSPKDYENTYQ 1158
 DB 181 YHAGKEHTSRKDMTAQNTGISTLYRNSYGAPADIKHNOVSAQPPQEPSPKDYENTYQ 240
 QY 1159 PFONSTRNYDSSFEDQVHRRPASEYTMHLGLKSTGVNDYFYSANRPSSELYSHYP 1218
 DB 241 PFONSTRNYDSSFEDQVHRRPASEYTMHLGLKSTGVNDYFYSANRPSSELYSHYP 300
 QY 1219 ASPDSMW 1225
 DB 301 ASPDSMW 307

RESULT 9
 Q9UPM3 PRELIMINARY; PRT; 114 AA.
 AC Q9UPM3;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE DELTA-CATENIN (FRAGMENT).
 OS Homo sapiens (Human).
 OC Chkaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Wilson, R., Smith, A., Elliott, G., Kramer, J., Latreille, P., Keppeler D.;
 RT "The sequence of H. sapiens BAC clone RG180P08A.";
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RA Waterston R.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 DR InterPro: IPR000225; -
 DR Pfam: PF00514; Armadillo_seg; 3.
 DR PROSITE: PS50176; ARM_REPEAT; 2.
 FT NON_TER 1 1
 SQ SEQUENCE 114 AA; 12961 MW; 66008C00A178D35D CRC64;

Query Match 9.3%; Score 114; DB 4; Length 114;
 Best Local Similarity 100.0%; Pred. No. 5.9e-107;
 Matches 114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 543 REFGRNDPELPEYIOMLOHOFPSVQSNAAAYLOHICFGDNKIKAEIRROGIOLVLDLD 602
 DB 1 REFGRNDPELPEYIOMLOHOFPSVQSNAAAYLOHICFGDNKIKAEIRROGIOLVLDLD 60
 QY 603 HRMTEVHRSACGLRLNLYGKANDDKIKALKNGGIPALVRLRTTDEIEL 656
 DB 61 HRMTEVHRSACGLRLNLYGKANDDKIKALKNGGIPALVRLRTTDEIEL 114

RESULT 10
 Q99569 PRELIMINARY; PRT; 1211 AA.
 AC Q99569;
 DT 01-MAY-1997 (TREMBLrel. 03, Created)
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE P0071 PROTEIN.
 OS Homo sapiens (Human).
 OC Chkaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=FRONTAL CORTEX;
 RX MEDLINE=97092329; PubMed=8937994;
 RA Hatfield M., Nachtsheim C.;
 RT "Cloning and characterization of a new armadillo family member, p0071, associated with the junctional plaque: evidence for a subfamily of closely related proteins."
 RL J. Cell Sci. 109:2767-2778(1996).
 DR EMBL; X81889; CAA57478.1; -
 DR InterPro: IPR000225; -
 DR Pfam: PF00514; Armadillo_seg; 7.
 DR PROSITE: PS50176; ARM_REPEAT; 3.
 DR SMART; SM00185; ARM; 1.
 SQ SEQUENCE 1211 AA; 134268 MW; 14C9626914A7024F CRC64;

Query Match 2.1%; Score 26; DB 4; Length 1211;
 Best Local Similarity 100.0%; Pred. No. 6.7e-17;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 913 ATALRNALDVNRKELIGRYAMRDLY 938
 DB 893 ATALRNALDVNRKELIGRYAMRDLY 918

RESULT 11
 Q9W5T9 PRELIMINARY; PRT; 779 AA.
 AC Q9W5T9;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE PUTATIVE P120CTN PROTEIN (CG17484 PROTEIN).
 GN P120CTN OR CG17484.
 OS Drosophila melanogaster (Fruit fly).
 OC Chkaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731137;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F., George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., Sutton G.G., Wortman J.R., Vandeil M.D., Zhang O., Chen L.X., Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D., Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G., Abtil J.F., Agbayani A., An H.-J., Andrews-Plannkoch C., Baldwin D., Ballew R.M., Basu A., Baxendale J., Bayraktaroglu T., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borikova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Canter A., Chandra I.,
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Deicher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Foster C., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
 RA Jajali M., Kalust F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., Moperson D.,
 RA Merkulov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Rhue B.C., Siden-Klamon I., Simpson M., Skupski M.P., Smith T.,
 RA Sier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RA "The genome sequence of *Drosophila melanogaster*."
 RA Science 287:2185-2195(2000).
 RA EMBL: AE002751; AAF5461.1; -
 RA HSSP: Q02248; 2BCT.
 RA FLYBase: FBgn0015587; p120ctn.
 RA InterPro: IPR000225; -
 DR Pfam: PF00514; Armadillo_seg; 4.
 DR PROSITE: PS50176; ARM_REPEAT; 3.
 DR SMART: SM00185; ARM; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 779 AA; 86588 MW; 5A1F546AE055CC2B CRC64;

Query Match 1.3%; Score 16; DB 5; Length 779;
 Best Local Similarity 100.0%; Pred. No. 6.3e-07;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 759 KTVENCVCILRNLSYR 774
 DB 397 KTVENCVCILRNLSYR 412

RESULT 12
 ID 09NHP1 PRELIMINARY; PRT; 781 AA.
 AC 09NHP1;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE ARM REPEAT PROTEIN.
 GN P120CTN OR CGI7484.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydriidae; Drosophilidae; Drosophila.
 NCBI_TaxId=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Cavallo R., Myster S., Peifer M.;
 RT "A *Drosophila melanogaster* homolog of the adherens junction protein
 p120ctn."
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF220496; AAF33245.1; -
 DR FLYBase: FBgn0015587; p120ctn.
 DR InterPro: IPR000225; -

DR Pfam: PF00514; Armadillo_seg; 5.
 DR PROSITE: PS50176; ARM_REPEAT; 3.
 DR SMART: SM00185; ARM; 1.
 SQ SEQUENCE 781 AA; 86836 MW; 5F8888EA349C7CEE CRC64;

Query Match 1.3%; Score 16; DB 5; Length 781;
 Best Local Similarity 100.0%; Pred. No. 6.3e-07;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 759 KTVENCVCILRNLSYR 774
 DB 429 KTVENCVCILRNLSYR 444

RESULT 13
 ID 095645 PRELIMINARY; PRT; 91 AA.
 AC 095645;
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
 DE PSILY2H-25 PROTEIN (FRAGMENT).
 GN PSILY2H-25.
 OS *Homo sapiens* (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Levesque G., Yu G., Fraser P.E., StGeorge-Hyslop P.;
 RT "Presentin 1 interacts with a novel protein which contains armadillo
 repeats and maps near the Cri du chat locus on chromosome 3p."
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U81005; AAD00454.1; -
 FT NON_TER
 SQ SEQUENCE 91 AA; 10606 MW; 06045C43A5CD807C CRC64;

Query Match 1.2%; Score 15; DB 4; Length 91;
 Best Local Similarity 100.0%; Pred. No. 1.1e-06;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 569 NAAAYLQHLCEGDNK 583
 DB 34 NAAAYLQHLCEGDNK 48

RESULT 14
 ID 09UP73 PRELIMINARY; PRT; 616 AA.
 AC 09UP73;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE P120 CATENIN ISOFORM 4AC.
 GN CTNND1.
 OS *Homo sapiens* (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE:96317528; PubMed:9653641;
 RA Keirsebilck A., Bonne S., Staes K., van Hengel J., Nollet F.,
 RA Reynolds A., Van Roy F.;
 RT "Molecular cloning of the human p120ctn catenin gene (CTNND1):
 expression of multiple alternatively spliced isoforms."
 RL Genomics 50:129-146(1998).
 DR EMBL: AF062318; AAC39803.1; -
 DR InterPro: IPR000225; -
 DR FLYBase: FBgn000308; -
 DR Pfam: PF00514; Armadillo_seg; 5.

DR Produm: PD000600; -: 1.
DR PROSITE: PSS0176; ARM_REPEAT; 3.
DR SMART: SM00185; ARM; 1
SQ SEQUENCE 616 AA; 68851 MW; E1014C183DAFB90D CRC64;

Query Match 0.9%; Score 11; DB 4; Length 616;
Best Local Similarity 100.0%; Pred. No. 0.06;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 568 SNAAYLQHLG 578
|||||
DB 61 SNAAYLQHLG 71

RESULT 15

Q9DE61 PRELIMINARY; PRT; 742 AA.
AC Q9DE61;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE P120 (FRAGMENT)
OS Xenopus laevis (African clawed frog)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Ciesiolka M., Vanlandschoot A., Staes K., van Roy F.;
RT "Armadillo-related proteins in Xenopus laevis."
RL Submitted (May-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF150746; AAG45945.1; -
FT NON_TER 1
SQ SEQUENCE 742 AA; 82927 MW; E51D3A3A5CB95FE7 CRC64;

Query Match 0.9%; Score 11; DB 13; Length 742;
Best Local Similarity 100.0%; Pred. No. 0.071;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 568 SNAAYLQHLG 578
|||||
DB 196 SNAAYLQHLG 206

See completed: July 19, 2001, 16:11:29
Job time: 218 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2001, 16:04:00 ; Search time 36.18 Seconds
(without alignments)
4479.648 Million cell updates/sec

Title: US-09-501-171A-4

Perfect score: 6377

Sequence: 1 MFARKPGGAFLGAMPVDPQ.....PYSELNETHYSPDPSWV 1225

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP-archaea:*
2: SP-bacteria:*
3: SP-fungi:*
4: SP-human:*
5: SP-invertebrate:*
6: SP-mammal:*
7: SP-mnc:*
8: SP-organella:*
9: SP-phage:*
10: SP-plant:*
11: SP-rodent:*
12: SP-unclassified:*
13: SP-vertebrate:*
14: SP-virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6377	100.0	1225	4	Q9UB33 homo sapien
2	6344	99.5	1225	4	Q00379 mus musculu
3	6129	96.1	1247	11	Q35927 mus musculu
4	5435	85.2	1040	4	Q9UM66 mus musculu
5	4598	72.1	876	4	Q43840 mus musculu
6	2818.5	44.2	1211	4	Q99569 mus musculu
7	1646.5	25.8	321	4	Q43206 mus musculu
8	1617	25.4	307	4	Q15390 mus musculu
9	1474.5	23.1	907	13	Q9DE92 xenopus lae
10	1459	22.9	742	13	Q9DE61 xenopus lae
11	1372	21.5	264	11	Q35116 rattus norv
12	1371	21.5	939	4	Q15088 mus musculu
13	1371	21.5	939	4	Q15088 mus musculu
14	1371	21.5	968	4	Q60716 mus musculu
15	1370	21.5	933	4	Q60935 mus musculu
16	1370	21.5	962	4	Q60713 mus musculu
17	1362	21.4	941	4	Q60713 mus musculu
18	1288	20.2	838	4	Q9UP71 mus musculu
19	1288	20.2	885	4	Q9UP72 mus musculu

20	1256	19.7	781	5	Q9NHP1	Q9NHP1 drosophila
21	1245.5	19.5	616	4	Q9UP73	Q9UP73 homo sapien
22	1171.5	18.4	779	5	Q9M579	Q9M579 drosophila
23	954.5	15.0	1181	5	Q9U308	Q9U308 caenorhabd
24	833	13.1	837	4	Q99960	Q99960 homo sapien
25	801	12.6	881	4	Q99959	Q99959 homo sapien
26	770.5	12.0	797	4	Q9Y446	Q9Y446 homo sapien
27	766	12.0	797	11	Q9OY23	Q9OY23 mus musculu
28	719.5	11.3	725	4	Q15152	Q15152 homo sapien
29	713	11.2	725	4	Q13835	Q13835 homo sapien
30	709.5	11.1	728	11	P97350	P97350 mus musculu
31	689	10.8	727	6	Q28161	Q28161 bos taurus
32	595	9.3	114	4	Q9UPW3	Q9UPW3 bos taurus
33	504	7.9	295	6	Q28875	Q28875 bos taurus
34	286	4.5	91	4	Q95645	Q95645 homo sapien
35	267.5	4.2	1566	4	Q9P286	Q9P286 homo sapien
36	266.5	4.2	995	4	Q9Y2W4	Q9Y2W4 homo sapien
37	265.5	4.2	1012	4	Q75359	Q75359 homo sapien
38	257	4.0	1190	4	Q99621	Q99621 homo sapien
39	256.5	4.0	1268	4	Q75046	Q75046 homo sapien
40	255.5	4.0	1012	4	Q43393	Q43393 homo sapien
41	255.5	4.0	1175	11	P70200	P70200 mus musculu
42	255.5	4.0	1182	4	Q99495	Q99495 homo sapien
43	245.5	3.8	1175	11	Q35126	Q35126 mus musculu
44	232.5	3.6	1006	11	Q62901	Q62901 rattus norv
45	224.5	3.5	3938	11	Q88778	Q88778 rattus norv

ALIGNMENTS

RESULT 1	PRELIMINARY:	PRT:	1225 AA.
ID Q9UB3			
AC Q9UB3			
DT 01-MAY-2000 (TREMBL) 13, Created			
DT 01-MAY-2000 (TREMBL) 13, Last sequence update			
DT 01-MAR-2001 (TREMBL) 16, Last annotation update			
DE NEURAL PLAKOPHILIN-RELATED ARM-REPEAT PROTEIN (NPRAP).			
OS Homo sapiens (Human).			
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
ON NCBI_TaxID=9606;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC TISSUE-BRAIN:			
RX MEDLINE-9223289; PubMed-10208590;			
RA Tanahashi H.;			
RT Isolation of human delta-catenin and its binding specificity with			
RL Neuroreport 10:563-568(1999).			
DR EMBL: AB013805; BAA36163.1; -			
DR InterPro: IPR000225;			
DR Pfam: PF00514; Armadillo_seg_6.			
DR PROSITE: P550176; ARM_REPEAT; 3.			
DR SMART: SM00185; ARM; 1.			
SO SEQUENCE 1225 AA; 132655 MW; 4A24ACC3E22BFE83 CRC64;			

2-8-1999

Query Match	100.0%	Score 6377;	DB 4;	Length 1225;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 1225;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY 1 MFARKPGGAFLGAMPVDPQSSASEKTSLSPLGNTSNGDSEETTTSAIIASVKEQEL 60				
Db 1 MFARKPGGAFLGAMPVDPQSSASEKTSLSPLGNTSNGDSEETTTSAIIASVKEQEL 60				
QY 61 QFERLTRELEAEROIVASQLEKCKLSGTSMSMSAEEQFOWQSDQGGKIDEDLTG 120				
Db 61 QFERLTRELEAEROIVASQLEKCKLSGTSMSMSAEEQFOWQSDQGGKIDEDLTG 120				
QY 121 LEVDSQIRSLQESGILDPDQYSGERPSLLSQSAIQLNKPGSGFOYPAVSNSQTLAL 180				
Db 121 LEVDSQIRSLQESGILDPDQYSGERPSLLSQSAIQLNKPGSGFOYPAVSNSQTLAL 180				

```

Db 121 LELVDSICIRSIQESGILDPDYSTGERPSLISQSLQINSKEGFSQYPASVHSMQTLAL 180
QY 181 GETTPSOILPARGTOARATGOSFSOGTTSRAGHLAGEPAPPPPPPPPPPPPPPSIGSAFHL 240
Db 181 GETTPSOILPARGTOARATGOSFSOGTTSRAGHLAGEPAPPPPPPPPPPPPPPSIGSAFHL 240
QY 241 PDAPPAALAAALYYSSSTLPAPPPRGSSPLAAPQGGSPPTKIQRGSAEGATYAAPRGSSP 300
Db 241 PDAPPAALAAALYYSSSTLPAPPPRGSSPLAAPQGGSPPTKIQRGSAEGATYAAPRGSSP 300
QY 301 KOSPSRLAKSYSTSSPINIVVSSAGLSPIRTVTPPYOOSTISSSPHQLSSTIGTYATLS 360
Db 301 KOSPSRLAKSYSTSSPINIVVSSAGLSPIRTVTPPYOOSTISSSPHQLSSTIGTYATLS 360
QY 361 PTKRLVHASEOYKSHSOELVATLQORGSILAAGSRASYSQHGHLGPBELALQSPHHI 420
Db 361 PTKRLVHASEOYKSHSOELVATLQORGSILAAGSRASYSQHGHLGPBELALQSPHHI 420
QY 421 DEYEDRVYQKPRMKSLSQSGDPLPAHNTGYRTSTAPSSPGVDVSVPLQRTGSOHGPN 480
Db 421 DEYEDRVYQKPRMKSLSQSGDPLPAHNTGYRTSTAPSSPGVDVSVPLQRTGSOHGPN 480
QY 481 AAAATFOAASIAAGPASYADPYRQLOQCPSESPYSKSGALPREGTLASPSIDSIQ 540
Db 481 AAAATFOAASIAAGPASYADPYRQLOQCPSESPYSKSGALPREGTLASPSIDSIQ 540
QY 541 DPREGMDPELPEYIOMLOHOPPSVOSNAAYLOHLCEFGDNKIKAEIRROGSIQLVLD 600
Db 541 DPREGMDPELPEYIOMLOHOPPSVOSNAAYLOHLCEFGDNKIKAEIRROGSIQLVLD 600
QY 601 LDHRTVEVHRSACGALRNLVYGKANDDNKIALKNCGGIPALVRLKRTTDEIRLTVG 660
Db 601 LDHRTVEVHRSACGALRNLVYGKANDDNKIALKNCGGIPALVRLKRTTDEIRLTVG 660
QY 661 LMNLSGCCALMPITIQDLAVLTANVITPHSGWENSPLODRKIQHSSQVLRNATGCLR 720
Db 661 LMNLSGCCALMPITIQDLAVLTANVITPHSGWENSPLODRKIQHSSQVLRNATGCLR 720
QY 721 NVSSAGEBARRMRCDELDTALVYIOSALGSEISDKTYENCVCILRNLSYLAETS 780
Db 721 NVSSAGEBARRMRCDELDTALVYIOSALGSEISDKTYENCVCILRNLSYLAETS 780
QY 781 OGOHMGTDLDGLCGEANGDAESSGCWKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 840
Db 781 OGOHMGTDLDGLCGEANGDAESSGCWKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 840
QY 841 LMHPSIVRPYITLSECSNPDTLEGAAGALQNLAAAGSWKWSVYIRAAVRKKGKGLPIVEL 900
Db 841 LMHPSIVRPYITLSECSNPDTLEGAAGALQNLAAAGSWKWSVYIRAAVRKKGKGLPIVEL 900
QY 901 LRINDRVYCAVATLRMALDVRNKLIGYARMDLVHRLPGGNNNSNNTASKAMSDTV 960
Db 901 LRINDRVYCAVATLRMALDVRNKLIGYARMDLVHRLPGGNNNSNNTASKAMSDTV 960
QY 961 TAVCCTLEHVITTKNNEKALRDAGIEKLVISKSGDKHSPPVYKAASOVLNMMOYR 1020
Db 961 TAVCCTLEHVITTKNNEKALRDAGIEKLVISKSGDKHSPPVYKAASOVLNMMOYR 1020
QY 1021 DURSILYKDDGMSOYHFAVSSSTIERDORPYSSTRTPSISVRYSPNNRSASAPSPREM 1080
Db 1021 DURSILYKDDGMSOYHFAVSSSTIERDORPYSSTRTPSISVRYSPNNRSASAPSPREM 1080
QY 1081 ISLKERKTDYECTGSNATYHGAKEHTSRKDAANTQNTGISTLYRNSYGAPAEIDIKNOV 1140
Db 1081 ISLKERKTDYECTGSNATYHGAKEHTSRKDAANTQNTGISTLYRNSYGAPAEIDIKNOV 1140
QY 1141 SAQVYQEPSRKDYETYPFONSTRANDSEFFEDQVHHRPPASEYTMHLGKSGNTYVD 1200
Db 1141 SAQVYQEPSRKDYETYPFONSTRANDSEFFEDQVHHRPPASEYTMHLGKSGNTYVD 1200
QY 1201 YSAARPYSELNETHSYSPASPSDW 1225
Db 1201 YSAARPYSELNETHSYSPASPSDW 1225

```

```

RESULT 2
ID 000379 PRELIMINARY; PRT: 1225 AA.
AC 000379; Q13589;
DT 01-JUL-1997 (TREMblrel. 04, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE DELTA-CATENIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
KA MEDLINE=97366296; PubMed=9223106;
RA Zhou J., Liyanage U., Medina M., Ho C., Simmons A.D., Kosik K.S.;
RT "Presentation of a novel member of the Armadillo family."
RL Neuroreport 8:2085-2090(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RA Lu Q., Paredes M., Medina M., Zhou J., Cavallo R., Peifer M.,
RA Orecchio L., Kosik K.S.;
RT "d-catenin, an adhesive junction associated protein which promotes
RT motile behavior."
RL Submitted (Oct-1998) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE OF 865-1225 FROM N.A.
RX MEDLINE=97202103; PubMed=9049630;
RA Simmons A.D., Overhauser J., Lovett M.;
RT "Isolation of cDNAs from the Cri-du-chat critical region by direct
RT screening of a chromosome 5-specific cDNA library."
RL Genome Res. 7:118-127(1997).
DR EMBL: 090197; FRC09409.1;
DR EMBL: 052828; AAB96357.1; -.
DR InterPro: IPR000225; -.
DR Pfam: PF00514; Armadillo_seg; 6.
DR PROSITE: PSS0176; ARM_REPEAT; 3.
DR SMART: SM00185; ARM; 1.
SQ SEQUENCE 1225 AA; 132664 MW; 8B676C8D5AFAE65 CRC64;

Query Match 99.5%; Score 6344; DB 4; Length 1225;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1220; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 361 PTKRLVHASEQYKSHQELVATATLQRPGLSLAASRASYSSQHGHLGPELRALQSPENHI 420
    |||
Db 361 PTKRLVHASEQYKSHQELVATATLQRPGLSLAASRASYSSQHGHLGPELRALQSPENHI 420
QY 421 DPTIEDRYVQKPPMSLSQSGDPLPRAHTGYRTSTAPSSPGVDSVPLQRTGSHGPN 480
    |||
Db 421 DPTIEDRYVQKPPMSLSQSGDPLPRAHTGYRTSTAPSSPGVDSVPLQRTGSHGPN 480
QY 481 AAAATFORASYAAGPASNADYRQLOQCPSESPYSKSGPALPEGTARSPIDSIOK 540
    |||
Db 481 AAAATFORASYAAGPASNADYRQLOQCPSESPYSKSGPALPEGTARSPIDSIOK 540
QY 541 DPREFGMRDPELPEVYIOMLQHPFVSQVSNAAAYLQHLCPGDKIKAEIRROGGIQLVYL 600
    |||
Db 541 DPREFGMRDPELPEVYIOMLQHPFVSQVSNAAAYLQHLCPGDKIKAEIRROGGIQLVYL 600
QY 541 DPREFGMRDPELPEVYIOMLQHPFVSQVSNAAAYLQHLCPGDKIKAEIRROGGIQLVYL 600
    |||
Db 541 DPREFGMRDPELPEVYIOMLQHPFVSQVSNAAAYLQHLCPGDKIKAEIRROGGIQLVYL 600
QY 601 LDHRTVEHRSACGALRNLYGKANDNKIALKNGGIPALVRLRKTDTLEIRELYGV 660
    |||
Db 601 LDHRTVEHRSACGALRNLYGKANDNKIALKNGGIPALVRLRKTDTLEIRELYGV 660
QY 661 LWNLSGCDALKNPITIDALAVLTNAVILPHSGWENSPLODDKRIOLHSSQVLRNATGCLR 720
    |||
Db 661 LWNLSGCDALKNPITIDALAVLTNAVILPHSGWENSPLODDKRIOLHSSQVLRNATGCLR 720
QY 721 NVSSAGEARRMRRCDCGLTDALLVYIOSALSSSEIDSKTVENCYILRNLSYRLAETS 780
    |||
Db 721 NVSSAGEARRMRRCDCGLTDALLVYIOSALSSSEIDSKTVENCYILRNLSYRLAETS 780
QY 781 QGQHMGTDELIDGLGCEANGKDAESSGCGKSKKKKKKKKKKKKKKKKKKKKKKKKKKK 840
    |||
Db 781 QGQHMGTDELIDGLGCEANGKDAESSGCGKSKKKKKKKKKKKKKKKKKKKKKKKKKKK 840
QY 841 LHMPSIVKPLTLTSSCSPDTEGAAGAALONLAAGSMKMSYIRAAVKEKGLPLVEL 900
    |||
Db 841 LHMPSIVKPLTLTSSCSPDTEGAAGAALONLAAGSMKMSYIRAAVKEKGLPLVEL 900
QY 901 LRIDNDRYVCAVATARNALADYRNKELIGKXAMRDLVHRLPGNNSNNTASKAMSDTY 960
    |||
Db 901 LRIDNDRYVCAVATARNALADYRNKELIGKXAMRDLVHRLPGNNSNNTASKAMSDTY 960
QY 961 TAVCCTLHEVITTKNMENAKALRDAGIEKLVGISKSGKSHPKVYKAAQVLYNSMOTR 1020
    |||
Db 961 TAVCCTLHEVITTKNMENAKALRDAGIEKLVGISKSGKSHPKVYKAAQVLYNSMOTR 1020
QY 1021 DLSLTKKQDMSQHYVYASSSTIERDORPYSSTPISPVVSPNNNSASAPASPRFM 1080
    |||
Db 1021 DLSLTKKQDMSQHYVYASSSTIERDORPYSSTPISPVVSPNNNSASAPASPRFM 1080
QY 1081 ISLKERKTDYECTGSNATYHGAKGERTSRKDAJTAONTGISTLYRNSYGAPADIKHNOV 1140
    |||
Db 1081 ISLKERKTDYECTGSNATYHGAKGERTSRKDAJTAONTGISTLYRNSYGAPADIKHNOV 1140
QY 1141 SAQPVQEPERKRYETVYOPQNTSTRYNDESEFFDOYVHNHPPASEYTMHLGKSTGYVDF 1200
    |||
Db 1141 SAQPVQEPERKRYETVYOPQNTSTRYNDESEFFDOYVHNHPPASEYTMHLGKSTGYVDF 1200
QY 1201 YSARPYSELYETESHYPASPSDMV 1225
    |||
Db 1201 YSARPYSELYETESHYPASPSDMV 1225

```

```

OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE=98002299; PubMed=9342840;
RA Pattenholz R., Franke W.W.;
RT "Identification and localization of a neutrally expressed member of the
RL plakoglobin/armadillo multigene family.";
DR EMBL: U90331; AAB82409.1;
DR MGI: 1195966; Ctnnd2.
DR InterPro: IPR000225;
DR Pfam: PF00514; Armadillo_seg; 6.
DR PROSITE: PS0176; ARM_REPEAT; 3.
DR SMART: SM00185; ARM; 1.
SQ SEQUENCE 1247 AA; 134998 MW; DA7A6B6A27D2919 CRC64;

Query Match 96.1%; Score 6129; DB 11; Length 1247;
Best Local Similarity 94.8%; Pred. No. 0;
Matches 1185; Conservative 8; Mismatches 29; Indels 28; Gaps 3;

QY 1 MPARKPEAAALGAMPVPDOPSSASEKTSLSPLGNTSNGDSTETTSALIASYKEQL 60
    |||
Db 1 MPARKQSGAAPFGAMPVPDOPPSASEKNSLSPLGNTSNGDSTETTSALIASYKEQL 60
QY 61 QPERLTRELAEROIIVASQLEKLGSEPTGSMSSSAEFOQWQODGOKDIEDELTTG 120
    |||
Db 61 QPERLTRELAEROIIVASQLEKLGSEPTGSMSSSAEFOQWQODGOKDIEDELTTG 120
QY 121 LEVYDSCIRSLQESGILDPDQYSTERPSSLSQALOLNSKEPSFOYPASYSNQTAL 180
    |||
Db 121 LEVYDSCIRSLQESGILDPDQYSTERPSSLSQALOLNSKEPSFOYPASYSNQTAL 180
QY 181 GETTSPQAPAGTOARATGSESGTTSRAGHLAGPEAPPPPPPPPPPPPPPPPPPPPP 240
    |||
Db 181 GDTASQPLASTQARRAQGSFSGTTRAGHLAGEPA-PPPPPPPPPPPPPPPPPPPP 239
QY 241 PDAPPAALAAALYSSSTPLPAPRPGSPPLAPOGGSPFKLQRGSAPEGATYAAPRGSSP 300
    |||
Db 241 PDAPP--AAALYSSSTPLPAPRPGSPPLTTOGGSPFKLQRGSAPEGATYAAPRGSSP 297
QY 301 KQSPRLAKSYSTSSPINIVVSSAGLSPIRVTSPPTVOSTISSPPIHOLSTICTATLS 360
    |||
Db 301 KQSPRLAKSYSTSSPINIVVSSAGLSPIRVTSPPTVOSTISSPPIHOLSTICTATLS 357
QY 361 PTKRLVHASEQYKSHQELVATATLQRPGLSLAASRASYSSQHGHLGPELRALQSPENHI 420
    |||
Db 361 PTKRLVHASEQYKSHQELVATATLQRPGLSLAASRASYSSQHGHLGPELRALQSPENHI 417
QY 421 DPTIEDRYVQKPPMSLSQSGDPLPRAHTGYRTSTAPSSPGVDSVPLQRTGSHGPN 480
    |||
Db 421 DPTIEDRYVQKPPMSLSQSGDPLPRAHTGYRTSTAPSSPGVDSVPLQRTGSHGPN 477
QY 481 AAAATFORASYAAGPASNADYRQLOQCPSESPYSKSGPALPEGTARSPIDSIOK 540
    |||
Db 481 AAAATFORASYAAGPASNADYRQLOQCPSESPYSKSGPALPEGTARSPIDSIOK 537
QY 541 DPREFGMRDPELPEVYIOMLQHPFVSQVSNAAAYLQHLCPGDKIKAEIRROGGIQLVYL 600
    |||
Db 541 DPREFGMRDPELPEVYIOMLQHPFVSQVSNAAAYLQHLCPGDKIKAEIRROGGIQLVYL 597
QY 598 DPREFGMRDPELPEVYIOMLQHPFVSQVSNAAAYLQHLCPGDKIKAEIRROGGIQLVYL 660
    |||
Db 598 DPREFGMRDPELPEVYIOMLQHPFVSQVSNAAAYLQHLCPGDKIKAEIRROGGIQLVYL 657
QY 601 LDHRTVEHRSACGALRNLYGKANDNKIALKNGGIPALVRLRKTDTLEIRELYGV 660
    |||
Db 601 LDHRTVEHRSACGALRNLYGKANDNKIALKNGGIPALVRLRKTDTLEIRELYGV 657
QY 661 LWNLSGCDALKNPITIDALAVLTNAVILPHSGWENSPLODDKRIOLHSSQVLRNATGCLR 720
    |||
Db 661 LWNLSGCDALKNPITIDALAVLTNAVILPHSGWENSPLODDKRIOLHSSQVLRNATGCLR 717
QY 721 NVSSAGEARRMRRCDCGLTDALLVYIOSALSSSEIDSKTVENCYILRNLSYRLAETS 780
    |||
Db 721 NVSSAGEARRMRRCDCGLTDALLVYIOSALSSSEIDSKTVENCYILRNLSYRLAETS 777

```


OY	781	OGOHNGTELDGILGCEANGKADASSGCMWKKKKKKSSODOMGVGLPPLCPACAPKGIOM	840
Db	778	OGOHNGTELDGILGCEANGKADTSSGCMWKKKKKKSSODOMGVGLPPLCPACAPKGIOM	837
OY	841	LMHPSIVAPYLLTSECSNPDTLEGAAGALONTAAGSMK-----	879
Db	838	LMHPSIVAPYLLTSECSNPDTLEGAAGALONTAAGSMKGMADVAGMAYALSLPEAG	897
OY	880	-----WSYIIRAAYKREKGLPIVLELLALINDRVCATALRMALDVRNKKELIGXYAMR	935
Db	898	CLPQWSSYIIRAAYKREKGLPIVLELLALINDRVCATALRMALDVRNKKELIGXYAMR	957
OY	936	DLVHRLPGGNNNSNNTASKAMDSDVTAVCCTLHEVLTAKNENAKALRDAGIEKLIGISK	995
Db	958	DLVHRLPGGNNNSNNSGSKAMSDDVTAVCCTLHEVLTAKNENAKALRDAGIEKLIGISK	1011
OY	996	SKGDGHSKPVYKKAASOVLNSMWOYRDLRSLYKKDGMWSOYHFVASSSTIERDORPYSSR	1051
Db	1018	SKGKHSKPVYKKAASOVLNSMWOYRDLRSLYKKDGMWSOYHFVASSSTIERDORPYSSR	1071
OY	1056	TPSISPVAVSPNNNSASAPASPREMISLTKERKTDYBECTGSNATYHAKGSHNSRKQAMTA	1111
Db	1078	TPSISPVAVSPNNNSASAPASPREMISLTKERKTDYBESAGNNATYHCKGSHNSRKQAMTA	1131
OY	1116	QNTGISTLYRNSYGAPEADIKHNDVSAQPOPEBSRKDYETVYOPFONSTRNNTYDESFEDQ	1171
Db	1138	QNTGISTLYRNSYGAPEADIKONQVSTQVPOPEBSRKDYETVYOPFONSTRNNTYDESFEDQ	1191
OY	1176	VHHRRPASEYTMHGLKSTGYNVDFYSAAPRYSELNETHSYHSPASDSWY	1225
Db	1198	VHHRRPASEYTMHGLKSTGYNVDFYSAAPRYSELNETHSYHSPASDSWY	1247

DE NEURAL PLAKOPHILIN RELATED ARM-REPEAT PROTEIN (FRAGMENT).
 GN NRPAP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RX MEDLINE=98002299; PubMed=9342840;
 RA Patenholz R., Franke W.W.;
 RT "Identification and localization of a neurally expressed member of the
 RT plakoglobin/armadillo multigene family.";
 RL Differentiation 61:293-304(1997).
 DR EMBL: U52351; AAB97957.1; -
 DR InterPro: IPR000225; -
 DR Pfam: PF00514; Armadillo_seg; 6.
 DR PROSITE: PS0176; ARM_REPEAT; 3.
 DR SMART: SM00185; ARM; 1.
 FT NON_TER
 FT SEQUENCE 876 AA; 96442 MW; 8D342B71C9E92A CRC64;

Query Match 72.1%; Score 4598; DB 4; Length 876;
 Best Local Similarity 99.9%; Pred. No. 5.3e-292;
 Matches 875; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 350 SSTIGYATLSTKRLVHASEQYKSHQELVATLQPGSLAASRAYSQHGHLPE 409
 DB 1 SSTIGYATLSTKRLVHASEQYKSHQELVATLQPGSLAASRAYSQHGHLPE 60
 QY 410 LRALQSPENHIDPIEDRYQKPKMSLSQSGDPLPAHGTGTYRTAPSSGVSVPL 469
 DB 61 LRALQSPENHIDPIEDRYQKPKMSLSQSGDPLPAHGTGTYRTAPSSGVSVPL 120
 QY 470 QRTGSHGQNAATAATFQASTAAGPASYADPYQDQYCEVESEPYKSGALPPEGL 529
 DB 121 QRTGSHGQNAATAATFQASTAAGPASYADPYQDQYCEVESEPYKSGALPPEGL 180
 QY 530 AASPSTDSIQKPRPFGMDPELPEYQMLQHPFVSQNAAYLOHCEFGNKKAETR 589
 DB 181 AASPSTDSIQKPRPFGMDPELPEYQMLQHPFVSQNAAYLOHCEFGNKKAETR 240
 QY 590 RGGIGDLDVLDLHRTVEHRSACGALRNLYGKANDNKIALKNCGGIPALVRLRKT 649
 DB 241 RGGIGDLDVLDLHRTVEHRSACGALRNLYGKANDNKIALKNCGGIPALVRLRKT 300
 QY 650 DLEIRLVGVLMNLSGSCALKPITIDALAVLTNAVITPHSGWENSPLODRKIQHSS 709
 DB 301 DLEIRLVGVLMNLSGSCALKPITIDALAVLTNAVITPHSGWENSPLODRKIQHSS 360
 QY 710 QVLRNATGCLRNVSSAGEARRRRCDDLTALVLYIOSALGSSSEIDSKTYENCYCLR 769
 DB 361 QVLRNATGCLRNVSSAGEARRRRCDDLTALVLYIOSALGSSSEIDSKTYENCYCLR 420
 QY 770 NLSYRLAETSGOQHMGTDLDLGCANGKDAESSGCGKRRKKSQDQDVGVPPL 829
 DB 421 NLSYRLAETSGOQHMGTDLDLGCANGKDAESSGCGKRRKKSQDQDVGVPPL 480
 QY 830 DCAEPKGIOMLHPSITVRYTLTSLSECSNPDTLEGACALONLAAGSKKSVYIRAAVR 889
 DB 481 DCAEPKGIOMLHPSITVRYTLTSLSECSNPDTLEGACALONLAAGSKKSVYIRAAVR 540
 QY 890 KEEGLPILVELLRINDRYVCAVATLRNALDVRKELIGYAMDLVHRLPGGNNSNN 949
 DB 541 KEEGLPILVELLRINDRYVCAVATLRNALDVRKELIGYAMDLVHRLPGGNNSNN 600
 QY 950 TASKASDDVTAVCTLHEVITRKNNENAKALRDAIGIEKLVGSKSKGDKSPKVVAA 1009
 DB 601 TASKASDDVTAVCTLHEVITRKNNENAKALRDAIGIEKLVGSKSKGDKSPKVVAA 660
 QY 1010 SOVLNWMQYRDLRSLYKKDQSHYFVASSSTIERDRQRPYSSSTPISIVRYSPNNR 1069

DB 661 SOVLNWMQYRDLRSLYKKDQSHYFVASSSTIERDRQRPYSSSTPISIVRYSPNNR 720
 QY 1070 SASAPASPREMISLKERKTDYECTGSNATYHGAKGHTSRKDMATQONTGISTLYRNSYG 1129
 DB 721 SASAPASPREMISLKERKTDYECTGSNATYHGAKGHTSRKDMATQONTGISTLYRNSYG 780
 QY 1130 APAEDIKHNQVSAQPVQPESRKDYETQPFONSTNRYDESEFEDOVHHRPASERTMIL 1189
 DB 781 APAEDIKHNQVSAQPVQPESRKDYETQPFONSTNRYDESEFEDOVHHRPASERTMIL 840
 QY 1190 GLKSTGYVDFYSAARYSLENTSHYPASDPDW 1225
 DB 841 GLKSTGYVDFYSAARYSLENTSHYPASDPDW 876

RESULT 6
 ID Q99569 PRELIMINARY; PRT; 1211 AA.
 AC Q99569;
 DT 01-MAY-1997 (Tremblrel. 03, Created)
 DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE P0071 PROTEIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=FRONTAL CORTEX;
 RX MEDLINE=97092329; PubMed=8937994;
 RA Hatzfeld M., Nachtsheim C.;
 RT "Cloning and characterization of a new armadillo family member, p0071,
 RT associated with the junctional plaque: evidence for a subfamily of
 RT closely related proteins.";
 RL J. Cell Sci. 109:2767-2778(1996).
 DR EMBL: X81889; CA57478.1; -
 DR InterPro: IPR000225; -
 DR Pfam: PF00514; Armadillo_seg; 7.
 DR PROSITE: PS0176; ARM_REPEAT; 3.
 DR SMART: SM00185; ARM; 1.
 DR SEQUENCE 1211 AA; 134268 MW; 14C9626914A7024F CRC64;

Query Match 44.2%; Score 2818.5; DB 4; Length 1211;
 Best Local Similarity 49.2%; Pred. No. 1.3e-175;
 Matches 618; Conservative 192; Mismatches 337; Indels 109; Gaps 37;

QY 15 MPVPDQSSASEKTSISLPGINTSNGSGSEFTT-SAILASVQEQLQRLRLRELEAER 73
 DB 1 MPVPDQSSASEKTSISLPGINTSNGSGSEFTT-SAILASVQEQLQRLRLRELEAER 59
 QY 74 QIVASOLEKRLGSETSGMSMSAEQFOFQMSOD-----GOKDIDELFTTGLLELVDS 127
 DB 60 QIVASOLEKRLGSETSGMSMSAEQFOFQMSOD-----GOKDIDELFTTGLLELVDS 115
 QY 128 IRSLOESGIL-DPODYSTERPDL--SOSALQJNSKEGSGFYQYASYSNQTALA----- 179
 DB 116 IRTEPEQGTLYSPQTSLSHESEGLSNGSRSTQNSYSDSGYQCFAGSFHNSQWYKADNR 175
 QY 180 -----LGTTSQSLPAKRTQARATGQSFQ-GTTSRAGHLAAGPPAPPP----- 224
 DB 176 QOHSFISGNTNHVY-----RNSRAEGQTLVQSVANRRMRRVSVPSRAQSPSYISTGVS 221
 QY 225 PPREPFAFLSAGFLP---DAPPAALAAALYSSSTLPAPPRGSPPLAPPOGSPTKLQ 281
 DB 232 PSKGLRTSLSGFSGSPVYDPRPLNPSA---YSSITLPA-ARASVY-SORASPTAIR 286
 QY 282 RGSNAPEGATYAADRGSSPK-QSPSRLAKSYSTSSPINIVSSAGSLPRLVTSPTVQST 340
 DB 287 RIGSVTSROT-SNPNGPPOYOTARV-----GSPILTIDAQF-----RVASPSQGO-V 333
 QY 341 ISSPFIHLSSTIGYATLSTP-KRLVHASEQYKSHQELVATLQPGSLAAGSASY 399

```

Db 334 GSSSP--KRSGMTAVPQIIGPSLQRTVHDMQFGQOQYDIERYVPRPSTL--TGLRSSY 330
QY 400 SSOGHGLPELRALQSPENHIDPIEDRYOKPMPRLSOSQDPLPAPITGYRTSTAP 439
Db 391 ASQSHQGLQDRLASVSPDLHTTPIYEGRTYYSPLYRSPNMGTYE--LOGSOTALYRT-- 445
QY 460 SSPGVDSV--PLQRTGSOHGPNNAATFORASYNAGPASNADYRLOJOCPSVESYSK 518
Db 446 ---GVSGIGNQRTSSOR-----STLYQNNVYALNTATAYAEYRIRYQRYVQ--ECYVNR 436
QY 519 SGPLPPEPGTILARSPSIDSTICKDPREFGMDPELPEVITOMLQHPSPVSGNAAAYLOHLC 578
Db 497 LQHAVPAPDGGTTRSPSIDSTICKDPREFAMRDPPELPEVITOMLQHPSPVSGNAAAYLOHLC 556
QY 579 FGDKKIAEIRROGILOLDLDHMRTEVYHRSAGALRLNLYVGKANDDKIKALNGGT 638
Db 557 FGDKKIAEIRROGILOLDLDHMRTEVYHRSAGALRLNLYVGKANDDKIKALNGGT 616
QY 639 PALVRLRKTDLERELVYGVLMNLSSCDALMPITQDALAVLTNAVITPHSGWENSPL 698
Db 617 PALVRLRKTDLERELVYGVLMNLSSCDALMPITQDALAVLTNAVITPHSGWENSPL 676
QY 699 QDDKIKIQLHSSQVLRNATGCLRNYSAGEBARRRMRECDGLDPLALYVIOALGSEIDS 738
Db 677 DDDKIKIQLHSSQVLRNATGCLRNYSAGEBARRRMRECDGLDPLALYVIOALGSEIDS 736
QY 759 KTVENCYILRLNLSYRLAETSGOGHGTDELGLCGEANGDAESSGCKKKKKKK 817
Db 737 KTVENCYILRLNLSYRLAETSGOGHGTDELGLCGEANGDAESSGCKKKKKKK 795
QY 818 --SODQMDGVLPDCAEPKGIOMLHPSIVRPLYTLSECSNPDTLEGAGALONLAA 875
Db 796 TPQEDQMDGVLPDCAEPKGIOMLHPSIVRPLYTLSECSNPDTLEGAGALONLAA 855
QY 876 GSKMKSYYTIAAAYKKEGLVLELRIDNDRYVCANATLRMALDYNKELIGYAMR 935
Db 856 GSKMKSYYTIAAAYKKEGLVLELRIDNDRYVCANATLRMALDYNKELIGYAMR 915
QY 936 DLVHRLPGNNSNTASKAMSDDTVTAVCTLEHVTIKNNENAKALRDAGIEKLGYSK 995
Db 916 DLVHRLPGNNSNTASKAMSDDTVTAVCTLEHVTIKNNENAKALRDAGIEKLGYSK 970
QY 996 SKGDKHSPKYYKAAQVLSNMQYRLRLSKYKKGDSQYHFAVSSSTIERDORPYSSR 1055
Db 971 GGRGRSLKYYKAAQVLSNMQYRLRLSKYKKGDSQYHFAVSSSTIERDORPYSSR 1026
QY 1056 TPSTISPV--RVSPNNRGSASAPAPREMISSKERTDYECGNSATYHGAKEGTSKDM 1113
Db 1027 TPSTISPV--RVSPNNRGSASAPAPREMISSKERTDYECGNSATYHGAKEGTSKDM 1083
QY 1114 TAONTGISTLYRNSYGAPAD---IKHNOVSAQVPOEPRSKDYETYOFPONSFRNYDE 1169
Db 1084 YPSSSKSPSYIYSSYPAREQNRRLQHOOLYSS--QODSNRKNFAYRLYLQSPHSYD 1141
QY 1170 SFPEDOVHHPRPASEYTMHGLKSTGNVYDFYGAARYSLANTYSHYSPSPSW 1225
Db 1142 PYFDVYHPR--PASTDYSTOYGLKSTGNVYDFYSTKRP---SYRABQYGSPPSW 1192

RESULT 7
043206 PRELIMINARY; PRT; 321 AA.
AC 043206;
DT 01-JUN-1998 (TREMblrel. 06, Created)
DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)
DE 01-MAY-2000 (TREMblrel. 13, Last annotation update)
DE PROBABLE CATEININ.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE=96207227; PubMed=8619474;
RA Andersson B., Wentland M.A., Ricatrente J.Y., Liu W., Gibbs R.A.;
RT "A 'double adaptor' method for improved shotgun library
RT construction."
RL Anal. Biochem. 236:107-113(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE=97264341; PubMed=9110174;
RA Yu W., Andersson B., Worley K.C., Muzny D.M., Ding Y., Liu W.,
RA Ricatrente J.Y., Wentland M.A., Lennon G., Gibbs R.A.;
RT "Large-scale concatenation cDNA sequencing."
RL Genome Res. 7:353-358(1997).
DR EMBL; AF035302; AAB8185.1;
KW Cytoskeleton; Structural protein; Cell adhesion.
SQ SEQUENCE 321 AA; 36108 MW; 1DDF4811891DA953 CRC64;

```

```

Query Match 25.8%; Score 1646.5; DB 4; Length 321;
Best Local Similarity 98.7%; Pred. No. 7.7e-100;
Matches 315; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 908 VVCA--VATALRMAALDVNKKELIGKYARDLVHRLPGGNSNNTASKAMSDTAVOCT 966
Db 3 VVCAVRSATLARMALDVNKKELIGKYARDLVHRLPGGNSNNTASKAMSDTAVOCT 62
QY 967 LHEVITKNNENAKALRDAGIEKLGYSKSKDKHSPKYYKAAQVLSNMQYRLRLSLY 1026
Db 63 LHEVITKNNENAKALRDAGIEKLGYSKSKDKHSPKYYKAAQVLSNMQYRLRLSLY 122
QY 1027 KKDGMQYHFAVSSSTIERDORPYSSSRTPSISVPRVSPNNRGSASAPAPREMISSK 1086
Db 123 KKDGMQYHFAVSSSTIERDORPYSSSRTPSISVPRVSPNNRGSASAPAPREMISSK 182
QY 1087 KTDVCTGSMATYHGAKEGHTSRKDMATONTGISTLYRNSYGAPADIEKHQVSAQVP 1146
Db 183 KTDVCTGSMATYHGAKEGHTSRKDMATONTGISTLYRNSYGAPADIEKHQVSAQVP 242
QY 1147 QPSPKDYETQOPQNSSTRNVDSEFFEDQVHHPRPASEYTMHGLKSTGNVDFYSAAR 1206
Db 243 QPSPKDYETQOPQNSSTRNVDSEFFEDQVHHPRPASEYTMHGLKSTGNVDFYSAAR 302
QY 1207 YSELNYETSHYPASPSW 1225
Db 303 YSELNYETSHYPASPSW 321

RESULT 8
015390 PRELIMINARY; PRT; 307 AA.
AC 015390;
DT 01-JAN-1998 (TREMblrel. 05, Created)
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
DE 01-AUG-1998 (TREMblrel. 07, Last annotation update)
DE G124.
GN G124.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX Fraser P.E., Levesque G., Rogaeva E.A., Yu G., St George-Hyslop P.H.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U72653; AAB6599.1;
SQ SEQUENCE 307 AA; 34417 MW; 35640DA46113B65F CRC64;

Query Match 25.4%; Score 1617; DB 4; Length 307;
Best Local Similarity 99.7%; Pred. No. 6.1e-98;

```


Query Match	21.5%;	Score 1372;	DB 11;	Length 264;
Best Local Similarity	100.0%;	Pred. No. 5e-82;		
Matches 264;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0
QY 563	FPVSQNAAYLQHLGCGDNKIKAEIRROGQIQLVLDLDRHMTVEVHRSAGALRLVYG 622			
DB 1	FPVSQNAAYLQHLGCGDNKIKAEIRROGQIQLVLDLDRHMTVEVHRSAGALRLVYG 60			
CY 623	KANDNNITALKNGGIPALVRLRKTDTLEIELVYGLVNLSSCDALKMPITD DALAVL 662			
DB 61	KANDNNITALKNGGIPALVRLRKTDTLEIELVYGLVNLSSCDALKMPITD DALAVL 120			
QY 683	TNAVITPHSGWNSPLQDDRKIQIQLHSSQYLNRNATGCLNNVSAGEEARRRMRECDGLTDA 742			
DB 121	TNAVITPHSGWNSPLQDDRKIQIQLHSSQYLNRNATGCLNNVSAGEEARRRMRECDGLTDA 180			
QY 743	LLVYIGALSSEIDSKTYEVCYCIIRNLSTYLAETSQGHGTDDELIGLGEANGKD 802			
DB 181	LLVYIGALSSEIDSKTYEVCYCIIRNLSTYLAETSQGHGTDDELIGLGEANGKD 240			
QY 803	AESSGCWKKKKKKKSODQWDYVG 826			
DB 241	AESSGCWKKKKKKKSODQWDYVG 264			
RESULT 12				
O60714	PRELIMINARY:	PRT:	939 AA.	
ID 060714				
AC 060714				
DT 01-AUG-1998 (TrEMBLrel. 07, Created)				
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)				
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)				
DE P120 CATEININ ISOFORM 1AC.				
GN CTNNDI.				
OS Homo sapiens (Human).				
OC Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;				
CC Mammalia; Eutheria; Primates; Catartini; Homiidae; Homo.				
OX NCBI_Taxid=9606;				
RA RN				
RA SEQUENCE FROM N.A.				
RP Kerszblak A., Bonne S., Staes K., Van Hengel J., Nollet F.,				
RA Reynolds A., Van Roy F.;				
RL Genomics 0:0-0(1998);				
DR EMBL: AF062321; AAC39806.1; -.				
DR InterPro: IPR000225; -.				
DR Pfam: PF00514; Armadillo_seg. 5.				
DR PROSITE: PS00176; ARM_REPEAT; 3.				
DR SMART: SM00185; ARM: 1				
SO SEQUENCE 939 AA; 104976 MW; D436D996BEBBDCT CRC64;				
Query Match	21.5%;	Score 1371;	DB 4;	Length 939;
Best Local Similarity	32.2%;	Pred. No. 3.2e-81;		
Matches 366;	Conservative 165;	Mismatches 322;	Indels 282;	Gaps 31;
QY 41	DSEETTTSAIIASVKEQELQERILTRLEAEKQIVASQLECKLG-SETGSMSSNSAE 99			
DB 2	DSEETTTSAIIASVKEQELQERILTRLEAEKQIVASQLECKLG-SETGSMSSNSAE 61			
QY 100	EQGWQSGQGDIEDLGTGLVDSQIRSLQSSGLDPQDYSTERPSLLSQSLQIN 159			
DB 62	RRHQNGRPVGDADLEKQKFSDKL-----NG---PDHS----- 92			
QY 160	SKPEGSFQVPASYHSNQTALGETTTPSQLPARGTQARATQSGFSGGTTNRAGHLGPEPA 219			


```

0Y 653 IRELTVGLVLMSCDMLKMP1I0DIALVATLVNVIITPHSGWEMSPLODORCKIQLHSSOYL 712
Db 468 LLEVITGLTLMLSHSDSKMEITVDHALHATLDEVIITPHSGWEPEDPCKPRHTEWYSL 527
0Y 713 RNATGCLNRVSSAGEEARRRMRRECGLTDLALLVYIOSALGSSSIDSKTEYNCVCILRNLS 772
Db 528 TINTAGCLTNVSSERSEARRRKLRCCDGLVADLFIWAQEIQRDSDKLVENCVCILRNLS 587
0Y 773 YRLAETISQCGHMGTDDELDDGLGCEANGKRAESSGCGWKKKKKKKKQDOW--DGYCPULD 830
Db 568 YQVHEIRLPQ-----AERYQEAAPRVANNTPGPHASOFGAKKGK---DMFSRGKKPIED 638
0Y 831 CA-----EPPKGIOMLMPHSIYKPYLTLLISCSNDTLEGAAGALOMLAAGSMK 880
Db 639 PANDTVDFPKRTSPARGELLFQGEVYRITISLKSCKPALILEAAGALIQMLCAGRWY 698
0Y 881 SVYIRAAVRKKKGCLPIVELLRIDNDRVCAVATALRNMALDVRENKELIGTYAMRDLYHR 940
Db 699 GRYIRLSALRQEKALISADLALTNHEHVEYRVAAAGALARNLAVARNNELIGKHAIRNLVKN 758
0Y 941 LPGGNSNNTKSKAMSDPYTAVOCTLHEVITRNMMNAKALRAGIEIKLVGSKSGDK 1000
Db 759 LPGGQN-----SNMFSDEYIYISLINTINETIAINLAPAKKIRETQGIETVLNLNS--GN 813
0Y 1001 HSPKVVKAASOVLYNSMWQYHDLBSLYKKDQMSQYHFVASSSTIENDR-ORPYSSSRPSI 1059
Db 814 RSEKEVRRAAALVDTQINGIKELRKLPELKEGWMKSDQVYNLNANASRQSSHSDSDTLPLI 873
0Y 1060 SPVAVSPNNRASAAPAPRPMISLKEKTDYECTGSNATYHGAKGHTSRKD 1111
Db 874 -----DRNQSDKKP--DREIQSNMGSMTKSLDNNYTPNREGHNTTLD 918

```

RESULT	14			
060716				
ID	060716	PRELIMINARY;	PRT;	968 AA.
AC	060716;			
DT	01-AUG-1998 (TREMBLrel. 07, Created)			
DT	01-AUG-1998 (TREMBLrel. 07, Last sequence update)			
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)			
DE	P120 CATENIN ISOFORMS 1ABC, 2ABC, 3ABC AND 4ABC.			
GN	CTNND1.			
OS	Homo sapiens (Human).			
OC	Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Keirenschick A., Bonne S., Staes K., Van Hengel J., Nollet F.,			
RA	Reynolds A., Van Roy F.;			
RL	Genomics 0:0-0(1998).			
DR	EMBL; AF062341; AAC33826.1; -			
DR	EMBL; AF062319; AAC33804.1; -			
DR	EMBL; AF062342; AAC33827.1; -			
DR	EMBL; AF062323; AAC33808.1; -			
DR	InterPro; IPR000225; -			
DR	Pfam; PF00514; Armadillo_seg; 5.			
DR	PROSITE; PS0176; ARM_REPEAT; 3.			
DR	SMART; SM00185; ARM; 1.			
SO	Alternative splicing.			
SO	SEQUENCE 968 AA; 108169 MW; D5C37489A891F292 CRC64;			

Query Match 21.5%; Score 1371; DB 4; Length 968;
Best Local Similarity 32.2%; Pred. No. 3.4e-81;
Matches 366; Conservative 165; Mismatches 322; Indels 282; Gaps 31.

[illegible]

Db	62	RRHNGRFVGADLDRKQFSDLK-----NG-----PQDHS-----	92
QY	160	SKRPSGFQYRASYHSNQTLALGCTTPPSQAPRKQTQAPATQOAFSQSFQCTTSRAGHLAPRPA	219
Db	93	-----	92
QY	220	PPPPPPPPFPFAPSLSAFHLPDAPAAAAALTYSSSTLPAPRPGSGPLAAPQGGSPRK	279
Db	93	-----HL-----LX-----STIP-----R	101
QY	280	LORGSAPRGATYAAPRGS-----SPKQSPRLAKSYSTSPINIVYSSA	324
Db	102	MOEPGQIYETIYEEDPEEAMSVSVETSDQGTTRRTTYKAKVKTWTTRQYVAMGPD	161
QY	325	GLSPIRVSPPTVOSTISSPILHDSST-----IGYATLSPTRKLIVHSDQY	372
..	162	GL-----PYDASSNNNTIQTGLGRDFKRNKGCGRPYVGAQATLPRNFHPDGT	213
QY	373	SKHSEIYATATLORGSIAAGSRASYSOHGLGP-----ELRALQSPENHIDPIYEDRY	428
Db	214	SHREYDGY-----PGG-----SDNYGSLSVTRYIEERYRPMQGYRAPNQD-V	256
QY	429	YQKPRMSLSQSGDPLRPATITGYRTSTAPSSGVSY-----PLQRTGSHGQ	479
Db	257	YQPOQYVRGSSVD-LHRFHPRYGLEDQSRMSGXUDLDLYGKMSDYGTARKRTGP----	311
QY	480	NAAATFORASYAAGPASNYADPKROYCPYSVPYSKSGP-----ALPRGTLA	530
Db	312	-----SDPRRRRL-----SYEDMIGEYVPDQYVYAPLQAHENGSLA	348
QY	531	RSPSIDSTQK-DPREFGMRDELPEVITQMLQHOPVYSVNAAYIQLHCFQDNKIKAEIR	589
Db	349	-----SDLSRKGGPPRPNMRQDELPEVITAMJGFRIDAVKSNAAAYIQLHCFQDNKIKAEIR	405
QY	590	ROGGIOLIVDLDHMHTEVHHSAGCALRNLYVGKANDKNTAKNCGGIRPALVRLIKRT	649
Db	406	KATGILPVVLGLDHPKKKVVHLAGAGALKNTSPGR-DODNKIATKNCDDGVAPVRLIKAR	464
QY	650	DIEIEELITGVLMNLSCGDALKMPTIODALAVLNTAAIYTHSGWENSPLODDRKIQLHSS	709
Db	465	DMDLFEVITGTLMLNLSHDSIKMETVDHALHALTDEVITPHSWEKEREPEDDCKPRIEME	524
QY	710	QVLRNATGCLRNVSAGGEARRRRREDDGLTDALLIYVQSALSSSTDSKYENKCYCIJR	769
Db	525	SVLTPTACLNRVNSERSEARRKLREDDGVDALEIFVQAEIQKDSKLIVENCYCLJR	584
QY	770	NLSYRLAETSQGHMGIDELDGLLCEANGKPAEISSCGMKKKKKKSDQDM--DGWCP	827
Db	585	NLSYVHHEIQ-----AERYQEAAPVAVANTGPHASCGAKKGR-----DEMFSRGGKP	635
QY	828	LPDCA-----EPKGIQMLMRPSIVKYPTLTLSECSNPDTELEGAALQNLAAIGS	877
Db	636	IEDPANDTVDEPKRTSPARGELLFQPEYVRIYISLTKESKTPALLIEMASGAIQMLCAGR	695
QY	878	WKMSYVYIAAARKKEGLPIIVETLIRJNDIVOCVAVATLARNALADVRENELLGKTYAMRL	937
Db	696	WTYGRYHSALRQEKALSAIDLLTNHEHYVKAASGALRNLAVDARNKEILLGKAIRPL	755
QY	938	VHRLPGGNSNNTASKAMSDTYTAVCCTLHEVITKMNENAKALRDAGIEKLIVGISRK	997
Db	756	VKNLPDGGQGN---SSWNFSEDTYISLNTINEYIAENLEAKKLRFQTEKTLVYLNKS-	811
QY	998	GDKHSPKYVKAASOVLNMOQYRLRSLYKKDQMSQYHNVYASSSTIERPR-QRPYSSST	1056
Db	812	GNRKSEKVRRAALVLOTINGYKELRPLEKEGWMKSDQVQMLNNAISRQSSHSTVDSSTL	870
QY	1057	PSISVPRVSPNNRBSASAPASPREMISLKEKRTDYECTGSNATYHGAKEHTSRKD	1111
Db	871	PLI-----DRNQKDKKP--DREIQMSNMGSNTKSLDNNYSTPRMERGHNTFLD	918

RESULT 15
060935

Fri Jul 20 10:09:45 2001

us-09-501-171a-4.rpt

Page 12

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2001, 16:00:10 ; Search time 27.68 Seconds

(without alignments)
2682.962 Million cell updates/sec

Title: us-09-501-171a-4

Perfect score: 6377

Sequence: 1 MFARKPPGAAPLGAMPVPO.....PSSELYETSHYPASPSDSW 1225

Scoring table:
BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A.Genesec 0601.*
1: /SID88/gcgdata/genesec/genesecp/AA1980.DAT.*
2: /SID88/gcgdata/genesec/genesecp/AA1981.DAT.*
3: /SID88/gcgdata/genesec/genesecp/AA1982.DAT.*
4: /SID88/gcgdata/genesec/genesecp/AA1983.DAT.*
5: /SID88/gcgdata/genesec/genesecp/AA1984.DAT.*
6: /SID88/gcgdata/genesec/genesecp/AA1985.DAT.*
7: /SID88/gcgdata/genesec/genesecp/AA1986.DAT.*
8: /SID88/gcgdata/genesec/genesecp/AA1987.DAT.*
9: /SID88/gcgdata/genesec/genesecp/AA1988.DAT.*
10: /SID88/gcgdata/genesec/genesecp/AA1989.DAT.*
11: /SID88/gcgdata/genesec/genesecp/AA1990.DAT.*
12: /SID88/gcgdata/genesec/genesecp/AA1991.DAT.*
13: /SID88/gcgdata/genesec/genesecp/AA1992.DAT.*
14: /SID88/gcgdata/genesec/genesecp/AA1993.DAT.*
15: /SID88/gcgdata/genesec/genesecp/AA1994.DAT.*
16: /SID88/gcgdata/genesec/genesecp/AA1995.DAT.*
17: /SID88/gcgdata/genesec/genesecp/AA1996.DAT.*
18: /SID88/gcgdata/genesec/genesecp/AA1997.DAT.*
19: /SID88/gcgdata/genesec/genesecp/AA1998.DAT.*
20: /SID88/gcgdata/genesec/genesecp/AA1999.DAT.*
21: /SID88/gcgdata/genesec/genesecp/AA2000.DAT.*
22: /SID88/gcgdata/genesec/genesecp/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6377	100.0	1225	21	AAB07973
2	6129	96.1	1247	21	AAB07974
3	5633	88.3	1084	20	AAV23900
4	5426	85.1	1040	18	AAW24559
5	3916	61.4	756	19	AAW06664
6	2818.5	44.2	1192	20	AAI23899
7	2818.5	44.2	1211	18	AAW24560
8	2818.5	44.2	1211	21	AAV23336
9	833	13.1	837	21	AAB36464
10	778.5	12.2	821	21	AAB36467
11	770.5	12.1	797	21	AAB36462

12	766	12.0	797	21	AAB36466
13	719.5	11.3	726	21	AAB36463
14	684	10.7	425	21	AAB33829
15	256.5	4.0	1185	20	AAV33497
16	255.5	4.0	1012	20	AAV17406
17	222	3.5	802	19	AAW37153
18	219.5	3.4	1837	21	AAI85564
19	218.5	3.4	787	19	AAW37152
20	218	3.4	787	19	AAW37151
21	194.5	3.1	2842	15	AAW33508
22	194	3.0	2843	16	AAW11922
23	194	3.0	2843	19	AAW76140
24	194	3.0	2843	19	AAW76144
25	194	3.0	2843	19	AAW33592
26	194	3.0	2973	19	AAW76821
27	194	3.0	2973	19	AAW76822
28	193.5	3.0	2385	21	AAW85569
29	192	3.0	2843	13	AAW26052
30	192	3.0	2843	18	AAW33592
31	192	3.0	2843	19	AAW38370
32	192	3.0	2860	15	AAW63507
33	191.5	3.0	781	21	AAW07290
34	191.5	3.0	781	21	AAW70740
35	190	3.0	1099	21	AAW70393
36	190	3.0	2973	21	AAW70304
37	189	3.0	1099	17	AAW05177
38	189	3.0	2971	21	AAW41231
39	188.5	3.0	2518	21	AAW40574
40	188	2.9	2972	22	AAW50363
41	188	2.9	3118	22	AAW50362
42	187	2.9	863	21	AAW42952
43	187	2.9	1130	17	AAW05178
44	187	2.9	1130	21	AAW70390
45	187	2.9	1130	21	AAW70394

ALIGNMENTS

RESULT 1	
ID	AAB07973
AA07973	standard; Protein; 1225 AA.
XX	
AC	AAB07973:
XX	
DT	14-NOV-2000 (first entry)
XX	
DE	A human neural plakophilin related armadillo protein.
XX	
KW	Human; presenilin; PS1; PS2; nerve cell growth; nerve damage; hnpRap;
KW	neural plakophilin related armadillo protein; Alzheimer's disease;
KW	Parkinson's disease; amyotrophic lateral sclerosis; ALS; stroke;
KW	multiple sclerosis; ischemia; stroke; neural paropathy; sciatic crush;
KW	motor neuron disease; peripheral neuropathy; neuropathy; diabetes;
KW	spinal cord injury; facial nerve crush.
XX	
OS	Homo sapiens.
XX	
PN	W0200047615-A2.
PD	17-AUG-2000.
XX	
PF	11-FEB-2000; 2000MO-CA00126.
XX	
PR	12-FEB-1999; 99US-0119835.
XX	
PA	(UTOR) UNIV TORONTO GOVERNING COUNCIL.
XX	
PI	St George-Hyslop PH, Fraser PE;
XX	
DR	WPI: 2000-524531/47.
XX	
DR	N-PSDB; AAA59700.
XX	

Mouse plakophilin-
Human plakophilin
Human cancer assoc
Human atrophin I p
Human atrophin-1
Mouse neutral Menat
Human homologue of
Mouse neutral Menat
Adenomatous polypo
Adenomatous polypo
Human APC protein
Human APC protein
Human APC protein
Transcriptional ac
Human homologue of
APC gene product i
Human adenomatous
Human adenomatous
Adenomatous polypo
Human beta catenin
Human beta catenin
Drosophila melanog
Protein used in ca
Lats large tumour
Human OREX ORE95
Human OREX ORE38
Human SRCAP. Homo
Human SRCAP. Homo
Human OREX ORE2716
H-Lats large tumou
Human Lats (large

2000 data

Gene

PT Stimulation of nerve cell growth using human Neural Plakophilin Related
 PT Armidillo protein (hnpRap) polypeptide, useful for the treatment of
 PT diseases such as Alzheimer's, Parkinson's, and stroke -
 PS
 XX Claim 3: Page 24-27; 33pp; English.

CC The present sequence represents human Neural Plakophilin Related
 CC Armidillo protein (hnpRap) polypeptide. hnpRap interacts with presenilins
 CC (PS1 and PS2) (PS1 and PS2). The specification describes a method
 CC for stimulating the growth of nerve cells, comprising contacting
 CC them with hnpRap. The hnpRap polypeptide and polynucleotide are useful
 CC for treating nerve damage caused by a variety of diseases or physical
 CC traumas, including Alzheimer's disease, Parkinson's disease, amyotrophic
 CC lateral sclerosis (ALS), multiple sclerosis, stroke, ischemia associated
 CC with stroke, neural paropathy, motor neuron diseases, sciatic crush,
 CC peripheral neuropathy, neuropathy associated with diabetes, spinal cord
 CC injuries and facial nerve crush.

CC Sequence 1225 AA:

Query Match 100.0%; Score 6377; DB 21; Length 1225;
 Best Local Similarity 100.0%; P-adj. No. 0;
 Matches 1225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFARKPGAAPLGAAMPVDPDSASEKTSISPGINTSGDSEETTSALIASVKDEEL 60
 DB 1 mfarpgpaaplgaampvdpdsasektsispgintsgdsettsaliasvkegel 60
 QY 61 QFERLTRELEAKQIVASQLEKRIKSETGSMSSMSAEQFOWOSODQKDEDELTTG 120
 DB 61 qferltreleaqivasaqlerkiqsetgsmssmsaeqfowosodqkdedelttg 120
 QY 121 LEIVNSCIRTSIOESIIIDPOYSGERSSILSOSALQINSEKSFQVPAHYNSQTLAL 180
 DB 121 leivnsctrsioesiiidpoysgerssilssosalqinseksfqpahynsqtlal 180
 QY 181 GETTPSQLPAGTQARATAGSOSFQGTTSRAGHLAGEPEAPPPPPRPFAPSLGSAFHL 240
 DB 181 gettpsqlpagtgatagsgsfqgttsragnlagpeappppprpfapslgsaahl 240
 QY 241 PDAPPAALAAALYSSSTLPAPRPGSGPLAPQGGSPKTLORGSGAPGATYAAPRGSSP 300
 DB 241 pdappaalaaalyssstlpaprpgrsgplapqggspktlorgsgapgatyaaprgssp 300
 QY 301 KOSPRFLAKSVSTSPINIVSSAGLSPIRTSPPTVOSTISSSPIHOLSTGTGATLS 360
 DB 301 kgsprflaksvstspiniwssaglsprtsptvostissspiholstgtgatlsl 360
 QY 361 PTKRLVHASSEYSKHSELYATATLQRPGLAAGSRASYSQHGHLGPELRALQSPENHI 420
 DB 361 ptkrlvhassseyskhselyatatlqrpgsllaagsrasysqhghlgpelralqspenhi 420
 QY 421 DPITYDRYQKPPMSSLSQSGDPLPRAHTGTTRTSPAPSPGVDSVPLQOTGSGOHQPN 480
 DB 421 dpytdryqkppmsslsqsgdplprahgttrtspapspgvdsvplqotgsgohqpn 480
 QY 481 AAAATFORASVYAGPASYADPYRLOVCPSESPYSKSGALPEGLASPSIDSTOK 540
 DB 481 aaaatforasvyaagpasyadpyrtlqycpsepsysksgalpeglaspsidsstok 540
 QY 541 DPREGMRDPELEVTOMLQHPFVSQSNAAAYLQHLFCGDNKIKAEIRROGGIQLVLDL 600
 DB 541 dpregmrdpelevtomlqhpfvsvsqsnaaaylqhlfcgdnkikaetirroggiqlvldl 600
 QY 601 LDHMTVEVHRSACGALRLVYTGKANDNKIALKNGGIPALVRLIKRTTDEIRLVYGV 660
 DB 601 ldhmtvevhrsacgalrlvlytgkandnkialknggipalvrlirkttdeirlyvgy 660
 QY 661 LWNSSCCALMPITQDLAVITNAVITPHSGWENSPLODDRKIQHSSQVLRNATGCLR 720
 DB 661 lwnssccalmpitqdlavltnavitphsgwensploddrikqhlssqvlrnatgclr 720

QY 721 NVSSAGEARRRRECDGLTDLALVYIOGALSSEIDSKTVENCYCILRNLSYRLAETS 780
 DB 721 nvssagearrrrrecdgltdlalyiogalsseidsktvenycilrnlstyrlaets 780
 QY 781 QGQHMGTDELIDGLGCEANGKDAESSGCWKKKKKKSSODQMDVGVLPCDAEPKGIQM 840
 DB 781 qgqhmgtdelidglgceangkdaessgcwkkkkkkssodqmdvgvlpdcapkgiqm 840
 QY 841 LHMPSIVKPYLTLLSCSNPDLEGAAGLQNLAAAGSWKWSYITAAARKKEGLITVEL 900
 DB 841 lhmpsi vkpylltllscsnpdllegaaaglnlaagswkwswyitaaarkkeglitvel 900
 QY 901 LRIDNRVVCATVALRNALDVKNELGKAMRDVRLPLGNNNSNNTASKAMSDTV 960
 DB 901 lriddnrvcavatalrimaldvkneligkaymrdvrlpgnnsnntaskamsdttv 960
 QY 961 TAVCGTLHEVYTKNMENAKALRDAGIEKLVGISKSGDKHSPKYKKAQVLSNMQYR 1020
 DB 961 tavcctlhevltkmenakalrdagieklygiskskgdkhspkykkaasqvlasmwy 1020
 QY 1021 DLRSLYKKDQMSQYHFASSSTIERDRORPSSRPTSPRYSPNNRSASAPSPREM 1080
 DB 1021 dlrslykkdgwsgyhtvaassstierdrprysstpslspvpspnnrsasapsrem 1080
 QY 1081 ISLKERKTDECTGSNATYHGAKGERTSRKDMTQNTGISTLYRNSYGAPAEIDIKHNOY 1140
 DB 1081 islkerktdectgsnatyhgakgertsrkdmtqntgistlyrnsygapaedikhnoy 1140
 QY 1141 SAQVPQEPKRDYETYPONSTRNYDSFEDQYHNPAPSETTMHGLKSTNYVDF 1200
 DB 1141 saqvpqepkrdyetyponstrnydsfedqyhnppasettmhglkstnyvdf 1200
 QY 1201 YSAARPYSELNETSHYPASPDGWSV 1225
 DB 1201 ysaarpyseletshypaspdgsw 1225

RESULT 2

AA07974
 ID AAB07974 standard; Protein; 1247 AA.

AC AAB07974;

DT 14-NOV-2000 (first entry)

DE A murine neural plakophilin related armidillo protein.

KW Human; presenilin; PS1; PS2; nerve cell growth; nerve damage; hnpRap;

KW neural plakophilin related armidillo protein; Alzheimer's disease;

KW Parkinson's disease; amyotrophic lateral sclerosis; ALS; stroke;

KW multiple sclerosis; ischemia; stroke; neural paropathy; sciatic crush;

KW motor neuron disease; peripheral neuropathy; neuropathy; diabetes;

KW spinal cord injury; facial nerve crush.

OS Mus sp.

PN WO2000047615-A2.

PD 17-AUG-2000.

PF 11-FEB-2000; 2000MO-CA00126.

PR 12-FEB-1999; 99US-0119835.

PA (UTOR) UNIV TORONTO GOVERNING COUNCIL.

PI St George-Hyslop PH, Fraser PE;

DR WPI; 2000-524531/47.
 DR N-PSDB; AAA59701.
 PT Stimulation of nerve cell growth using human Neural Plakophilin Related
 PT Armidillo Protein (hnpRap) polypeptide, useful for the treatment of

PT diseases such as Alzheimer's, Parkinson's, and stroke -

XX PS Disclosure; Page 30-33; 33pp; English.

XX The present sequence represents Neural Plakophilin Related Armadillo Protein (NPRAP) polypeptide. Human NPRAP interacts with presenilins (PS) I and II (PS1 and PS2). The specification describes a method for stimulating the growth of nerve cells, comprising contacting them with hNPRAP. The hNPRAP polypeptide and polynucleotide are useful for treating nerve damage caused by a variety of diseases or physical traumas, including Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis (ALS), multiple sclerosis, stroke, ischemia associated with stroke, neural atrophy, motor neuron diseases, sciatic crush, peripheral neuropathy, neuropathy associated with diabetes, spinal cord injuries and facial nerve crush.

XX Sequence 1247 AA;

Query Match 96.1%; Score 6129; DB 21; Length 1247;
Best Local Similarity 94.8%; Pred. No. 0;
Matches 1185; Conservative 8; Mismatches 29; Indels 28; Gaps 3;

QY 1 MFARPCGAPLIGAMPVDDPSSASEKTSLSPLNTSNGDSEFTTSAIIASVKEQL 60
DB 1 mfarpcgaaplignampvddqppasaseknsslsplntnsngdsettsailasvkeql 60

QY 61 QFERLTRELEAERQIVASQLEKRCIKGSETGSMSSAEEQPMQSDGQKDEDELTTG 120
DB 61 qferltreleaeqrlvasqlercklgetsgsmssaaeeqpmqsdgqkdedelttg 120

QY 121 LEIVDSCTRSIQESGILDPDOSTGERPSSLISQALQNSKPGESFOYASYSNOTLAL 180
DB 121 leivdsctrsiqesgilddpdostgerpsllsqalqnskgesfoyasysnotlal 180

QY 181 GETTPSQLPARGTQARATGSEFSGTSTRAGHLAGPEPAPPPPPPPPPPPPPPPPSAFHL 240
DB 181 gettpsqlpargtqaratgsefsgtstraghlagpepapppppppppppppppsafhl 240

QY 241 PDAPPAALAAALYSSSTPLAPPPGSSPLAAPGGSPTKLORSGSAPBETATYAAPRGSSP 300
DB 241 pdapp--aaalyssstlpppprgssplttggsptklorgsgsapbetatyaaargsssp 300

QY 301 KQSPRLAKSYSTSPNTIVSSAGLSPTRVTSPPVOSTISSPTIHQSTTGTATLS 360
DB 301 kqspriaksyssntivssaglsptrvtsppvostissptihqsttgtatls 360

QY 361 PTKRLVHASBOYSKHQELATATLQRPCSLAAGSRASYSQGHGHPRLALQSPENHI 420
DB 361 ptkrlvhasboyskhqelatatlqrpsslaagrsasysqghhlpelalqspenh 420

QY 421 DPTIEDRVYQKPRKSLSSQSGDPLPRAHTGTYRTSTABSPGVSVPLORTSGHGPON 480
DB 421 dptiedrvyqkprkslssosgdplprahtgtyrtstabspgvsvplortsgohpon 480

QY 481 AAAATFORASYAAGPANTADPYROLOVPSVESPYSKSGAPLPEBGLTANSPSIDSIQK 540
DB 481 aaaatforasyaagpantadpyrtdycaavspysksgaplpebgltarpsidsiqk 540

QY 541 DPREGWRDELEPEVLOLQHPFSVSNAAAYLOHLFCGDKRIKAEIRROGGIOLVLDL 600
DB 541 dpregrwdelepevliqlqhpfsvsnaaaylqhlfcgdkrikaeirrroggiolvldl 600

QY 601 LDHRTMEVRSAGALRNLYYKANDDKKIALKCGGIPALVRLIKRTTDELEIRELYTGV 660
DB 601 ldhrtmevrsagalnlyykanddkialkcggipalvrlirkrttdeleirelytgv 660

QY 661 LWNLSGCDALKMPITODALAVLTNAVITPHSGWENSPLODDRKIOHLSOVLRAATGCLR 720
DB 661 lwnlsgcdalkmpitodalavltnavitphsgwensploddrkiohlssovlraatgclr 720

QY 721 NVSSAGEEARRMRRCDCGTLTDLALVYIOSALGSSEIDSKTVENCVCILRNLSYLAETS 780
DB 721 nvssageearrmrrcdbgltldalvyiosalgsseidsktvencvcilrnlsylaets 780

DB 718 NVSSAGEEARRMRRCDCGTLTDLALVYIOSALGSSEIDSKTVENCVCILRNLSYLAETS 777

QY 781 OGGHMTDELDELICGEANGKDAESSGCMGKKKKKKKSGODGQVGLPDCAPKPKIOM 840
DB 778 qggmgtdelldlclgeangkdaessgcwkkkkkksgdqgdvgvplpdcapppkqigm 837

QY 841 LWHPSLVKRYTLTLSECSNPDTLEGAGALQNLAAAGSNK----- 879
DB 838 lwhpslvkrytlltlsecsnpdtlegagalqnlagswkyaedvagmayalslpegap 897

QY 880 ----SVYFRAAVREKELPLIVELRLDNRVVCAYATALRNALDVRNKEILGYAMR 935
DB 898 clpqsavvyrfaavrekelpilvelrlidndrvvcavatalrimaldvrynkeilgyamr 957

QY 936 DLVHRLPGGNNSNNTASKMSDDVTAVCCYLHEVIRKNNENKALPDAGITKLVGISK 995
DB 958 dlvrhrlpggnnsnntasgkmsddvtavccylhevlrknnenaalrdaagilekvlgisk 1017

QY 996 SKGDHSHKRYKAAQVLNSWQYRDLRLSLYKKDQWQYHFVASSSTIERDORPYSSSR 1055
DB 1018 skgdhshkrykvaasqvlnswwyrdlrsllykkdqwgyhfvasstierdorypsssr 1077

QY 1056 TPISIPVTVSPNNNSASAPASPREMISLKERKNDYECTGSNATYHGAKGPHSRKDAMTA 1115
DB 1078 tpsisptvtpnnnsasapapremislkerkndyectgsnatyhgakgphsrkdamta 1137

QY 1116 QNTGISLTYRNSYAPAEKDKHNOVSAQVPOBPSRKDYETQYPOFNSTRNDESEFEO 1175
DB 1138 qntgisltyrnsyapaeedkqngvstqpyppgsrktqdyetqyfnstrndesefedq 1197

QY 1176 VHHRRPASEYTMHGLKSTGNYDFYSAAAPYSELNFTSHYPASDSWV 1225
DB 1198 vhhrrpaseytmhlykstgnydfysaaapyselnfetshypasdsww 1247

RESULT 3
AAV23900
ID AAV23900 standard; protein; 1084 AA.
XX
AC AAV23900;
DT 27-SEP-1999 (first entry)
DE Human resenilin binding armadillo protein GT24/hNPRAP.
XX
KW Human; presenilin 1; PS1; presenilin-binding protein; interacting domain;
KW presenilin allele; Alzheimer's disease; senile dementia;
KW psychiatric disease; schizophrenia; depression; neurological disease;
KW stroke; cerebral haemorrhage; p0071; armadillo protein.
XX
OS Homo sapiens.
XX
FH
FT Key Location/Qualifiers
FT Misc-difference 115 /note= "not specified"
PN MO9935501-A1.
PD 15-JUL-1999.
PE
PF 08-JAN-1999; 99WO-CA00018.
PR 09-JAN-1998; 98US-0070948.
PA (UTOR) UNIV TORONTO GOVERNING COUNCIL.
XX
PI Fraser PE, St George-Hyslop PH;
XX WPI, 1999-419410/35.
PT Identifying substances that alter presenilin interactions, useful
PT for screening individuals for presenilin alleles associated with
PT Alzheimer's disease - useful for diagnosis of Alzheimer's disease

XX Claim 9; Fig 1; 40pp; English.
 PS The present sequence represents a human presenilin-binding protein,
 CC termed GR24/human neuronal plakoglobin related armadillo protein (GR24/
 CC hNRPAP). The specification describes a method for identifying substances
 CC that alter the interaction of a presenilin with a presenilin-binding
 CC protein. The method comprises contacting the interacting domain of a
 CC presenilin protein to a presenilin-binding protein in the presence of
 CC a test substance, and measuring the interaction of the presenilin and
 CC the presenilin-binding protein. The method can be used to screen
 CC individuals for presenilin alleles associated with Alzheimer's disease
 CC and related disorders, such as senile dementia's, psychiatric diseases
 CC such as schizophrenia and depression, and neurological disease, such as
 CC stroke and cerebral haemorrhage.
 CC
 CC Sequence 1084 AA;
 SQ
 Query Match 88.3%; Score 5633; DB 20; Length 1084;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1076; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 148 PSLTSSALQNSKPEGSFOYDASYSNOTLALGETTPEQLPARGTOARATGOSFSOGTT 207
 7 PSLTSSALQNSKPEGSFOYDASYSNOTLALGETTPEQLPARGTOARATGOSFSOGTT 66
 208 SRAGHLAGEPEAPPP 267
 67 SRAGHLAGEPEAPPP 126
 268 PLAAPOGGSTKIQKGSAPGATYAPRGSPPKSPKSPKSPKSPKSPKSPKSPKSPKSP 327
 127 PLAAPOGGSTKIQKGSAPGATYAPRGSPPKSPKSPKSPKSPKSPKSPKSPKSPKSP 186
 328 PIRVTSPPVOSTISSPSSIHQLSTTIGYATLSPTRKLVHASEQYSKHSOELVATTLOR 387
 187 PIRVTSPPVOSTISSPSSIHQLSTTIGYATLSPTRKLVHASEQYSKHSOELVATTLOR 246
 388 PGSLAAGSRAYSOSQGHGCLPELRALQSPHNTIDRYEDRYOKPMPRSLSOSQGDPLPP 447
 247 PGSLAAGSRAYSOSQGHGCLPELRALQSPHNTIDRYEDRYOKPMPRSLSOSQGDPLPP 306
 448 AHNTGYRTAPSSPGVDVYPLQRTGSOHGPONAAAATFORASYAAGPASNVADPYRQIQ 507
 307 AHNTGYRTAPSSPGVDVYPLQRTGSOHGPONAAAATFORASYAAGPASNVADPYRQIQ 366
 508 YCPVESPYSKSGPALPPEGLTARSPSIDSIQKDPREFGMRDELPLVYIOMLOHOFPSVQ 567
 367 YCPVESPYSKSGPALPPEGLTARSPSIDSIQKDPREFGMRDELPLVYIOMLOHOFPSVQ 426
 568 SNAAYIQLHCFQGNKTKAETIRROGIGIQLVDLDHMTFVHRSAGCALNRLVYKANDD 627
 427 SNAAYIQLHCFQGNKTKAETIRROGIGIQLVDLDHMTFVHRSAGCALNRLVYKANDD 486
 628 NKIALKNGGIPALVRLKRTTDLERLVTGVLMNLSGCCALMPTIIDAALVLTNAVY 687
 487 NKIALKNGGIPALVRLKRTTDLERLVTGVLMNLSGCCALMPTIIDAALVLTNAVY 546
 688 IPHSGWENSPLODDRKTIOHSSOYLKRNATGCLRWVSSAGFEARRRRCOGITPALLVYI 747
 547 IPHSGWENSPLODDRKTIOHSSOYLKRNATGCLRWVSSAGFEARRRRCOGITPALLVYI 606
 748 OSALGSSRIDKTYENCVCILRLSYRLAETESOGHGTDELGLLGEANGKADASSG 807
 607 OSALGSSRIDKTYENCVCILRLSYRLAETESOGHGTDELGLLGEANGKADASSG 666
 808 CWGKK 867
 667 CWGKK 726
 868 GALQNLAAAGSMWWSYIIRAARKEKGLPLIVELLRIIDDRVYCAVATLRNNALDVRNKE 927

DB 727 galqnlaaagsmwswyilaarvkekgjpllvellriidndrvycaavatalrnnaldrvnrke 786
 QY 928 LICKYAMRDLVHRLPGGNSNNNTASKAMSDPTVAVCCITHEVITKNNENAKALDAGGI 987
 DB 787 LICKYAMRDLVHRLPGGNSNNNTASKAMSDPTVAVCCITHEVITKNNENAKALDAGGI 846
 QY 988 EKIYVLSKSGKDNHSPKVVYKASQVLYNSMWORYDLNLSLYKKDQMSQYHFVASSSTIENDR 1047
 DB 847 EKIYVLSKSGKDNHSPKVVYKASQVLYNSMWORYDLNLSLYKKDQMSQYHFVASSSTIENDR 906
 QY 1048 QRYSSSRTPSISIPVSPRNRSASAPSPREMLSKERTYECTGNSNATYHAKAGERT 1107
 DB 907 QRYSSSRTPSISIPVSPRNRSASAPSPREMLSKERTYECTGNSNATYHAKAGERT 966
 QY 1108 SRKDAWNTAGTISTLYRNSYGAPADIRKNOVSAQVPQEDSRKDYETTYOPONSTRNY 1167
 DB 967 SRKDAWNTAGTISTLYRNSYGAPADIRKNOVSAQVPQEDSRKDYETTYOPONSTRNY 1026
 QY 1168 DESFEEDQVHHRPPASITMHLGLKSTGNYVDFYSAARPSSELYETSHYPAASPSMV 1225
 DB 1027 DESFEEDQVHHRPPASITMHLGLKSTGNYVDFYSAARPSSELYETSHYPAASPSMV 1084
 RESULT 4
 AAW24559 standard; Protein: 1040 AA.
 ID AAW24559;
 AC AAW24559;
 XX 06-FEB-1998 (first entry)
 DE Presenilin-Interacting protein GT24.
 KW Presenilin-Interacting protein; human; Alzheimer's disease;
 KW diagnosis; therapy; transgenic animal; animal model; GT24.
 XX Homo sapiens.
 XX
 XX Key Location/Qualifiers
 XX Domain 346..862
 XX FT /note= "presenilin-interacting domain"
 XX FT Misc-difference 12
 XX FT /note= "encoded by GST"
 XX FT Misc-difference 40
 XX FT /note= "encoded by CCR"
 XX FT Misc-difference 71
 XX FT /note= "encoded by AKC"
 XX
 XX W09727296-A1.
 XX 31-JUL-1997.
 XX 27-JAN-1997; 97WO-CA00051.
 XX 02-JAN-1997; 97US-0034590.
 XX 26-JAN-1996; 96US-0592541.
 XX 05-JUL-1996; 96US-0021673.
 XX 12-JUL-1996; 96US-0021700.
 XX 08-NOV-1996; 96US-0029895.
 XX
 XX (HSCR-) HSC RES & DEV LP.
 XX (UTOR) UNIV TORONTO GOVERNING COUNCIL.
 XX
 XX Frazer PE, Rommens JM, St George-Hyslop PH;
 XX WPI: 1997-393684/36.
 XX N-PSDB: AAT79967.
 XX
 XX Presenilin-interacting protein genes - used to develop products for
 XX the diagnosis, therapy and study of Alzheimer's disease and related
 XX disorders
 XX
 XX Claim 1; Page 99-101; 133pp; English.


```

Db 228 gldeapeatevpsrtlr-apaavrtlqrfqsgshrsrgvgavpavlepr---varapavrs 283
OY 538 IQ-----KDPREF-----GMRDPELPEYIOMLQHOFPSVOSNAAYL 574
Db 284 lslsladaghlphdyhtnsysshrtlqrlssgfdiddlpsavkylmsdphlqylgaayl 343
OY 575 QHLCEGDNKIKAEIRROGGIOLVLDLDRHMTVEHRSACGALRNLYVGKANDDKIKALKN 634
Db 344 qhkcydaaakqgslqavprlvklfnhanqevqratgamrnllyd--nadrkklalve 401
OY 635 CGGIPALVRLRKTTDTLEIRLVTGVLNLSGCDLKPITQDALAVLNAVITPHSGWE 694
Db 402 enqifellrtlreqedd-elrknavgllwnlsssdhklrdlartleqldtlvlspsag 460
OY 695 NSPIODDRKILHSS--OVLNATGCIERNSSAGEARRRMREDCGLTDLALVYIOSALG 752
Db 461 gppl-----lqnasaeafynatvgflrnlssasqatrqkmrtchglvdaivslnhal 515
OY 753 SSEIDSKTVENCVCILRNLSYRL-----AETSOGHMGTELDLGLCEANGKDAES 805
Db 516 agkedksvenavcvinlsyrllydempesalqrlqgr--gridlagappegv----- 566
OY 806 SGCGKSKKKKKSSODQDVGPRDCAEPKGIOMLHNPSTVAPYLTLESCS-NPDTLE 864
Db 567 vgcftpgsrrlreiplaadaltfaevskdpqglewlspsqlvlylrlqrcelnhtle 626
OY 865 GAAGALONLAGSWKWS-VYTRAVRREKCPILVELLRIDNDVCAVATLRNALD 923
Db 627 aaagalnigtadgrtwagvlsrlalegerllnplldvrtcadhqlrlslgllnlsrna 686
OY 924 RNKELICKYAMRDLVHRLPGNNNSNTASKAMSDDTVAVCCTLHEVITNMENAKALRD 983
Db 687 rnkdemstkvshlklep-----svgekspaeavlwnlavllnlvasplaaardlly 741
OY 984 AGGIEKLVGISKSGDKHSPVYKASOVLSNMWYRDLSIKKQDSQYHMY 1037
Db 742 fdgrlrkliflkkrrdspsekssraassllanlwgynklhndirakxyrkedil 795

RESULT 12
AAB36466
ID AAB36466 standard; Protein; 797 AA.
AC AAB36466;
XX
Df 01-MAR-2001 (first entry)
XX
DE Mouse plakophilin-3 protein SEQ ID NO:7.
XX
KW Plakophilin-3; PKP3; catenin-like protein; characterisation; diagnosis;
KW desmosome; epithelial cell; skin disease; dermatological; gene therapy;
KW vaccine.
XX
OS Mus musculus.
XX
WO200006619-A2.
XX
PD 09-NOV-2000.
XX
PP 28-APR-2000; 2000MO-EP04389.
XX
PR 30-APR-1999; 99EP-0870093.
XX
PA (VLAAR-) VLAAHS INTERUNIVERSITAIR INST BIOTECHNOG.
XX
PI Van Roy F, Bonne S;
XX
DR MPI; 2000-687529/67.
XX
N-PSDB; AAC64793.
XX
PT Nucleic acids encoding Plakophilin-3 polypeptides, useful for treating
PT skin diseases and disorders of epithelial tissue associated with

```

```

PT inappropriate Plakophilin-3 expression and activity -
XX Disclosure: Fig 11; 132pp; English.
XX
XX The present invention describes an isolated or recombinant nucleic acid
XX molecule (I) encoding a Plakophilin-3 (PKP3), from humans, mice and
XX Xenopus laevis. (I) has dermatological activity, and can be used in
XX gene therapy and for vaccines. (I) and the protein it encodes may be
XX used in the prevention, treatment and diagnosis of diseases associated
XX with inappropriate PKP3 expression, such as skin diseases and disorders
XX affecting epithelial tissue. For example, (I) (and vectors containing
XX (I)) and the PKP3 polypeptide may be used to treat disorders associated
XX with decreased PKP3 expression by rectifying mutations or deletions in
XX a patient's genome that affect the activity of PKP3 by expressing
XX inactive proteins or to supplement the patients own production of PKP3
XX polypeptides. Additionally, (I) may be used to produce PKP3, according
XX to standard recombinant DNA methodology, by inserting the nucleic acids
XX into a host cell and culturing the cell to express the protein. (I) and
XX complementary sequences may also be used as DNA probes in diagnostic
XX assays to detect and quantitate the presence of similar nucleic acid
XX sequences in samples, and hence which patients may be in need of
XX restorative therapy. The PKP3 polypeptides may be used as antigens in
XX the production of antibodies against PKP3 and in assays to identify
XX modulators (agonists and antagonists) of PKP3 expression and activity.
XX The anti-PKP3 antibodies and PKP3 antagonists may also be used to down
XX regulate PKP3 expression and activity. The anti-PKP3 antibodies may also
XX be used as diagnostic agents for detecting the presence of PKP3
XX polypeptides in samples (e.g. by enzyme linked immunosorbant assay
XX (ELISA)). PKP3 is a catenin-like protein, which is present in desmosomes
XX and nuclei of epithelial cells. The present sequence represents mouse
XX PKP3, as given in the present invention.
XX
XX Sequence 797 AA;
XX
XX Query Match 12.0%; Score 766; DB 21; Length 797;
XX Best Local Similarity 27.5%; Pred. No. 3,66-46;
XX Matches 266; Conservative 125; Mismatches 380; Indels 196; Gaps 26;
OY 85 LSGETGSMSSMSAEQFOWQSGDKDIEDELITGGLVDSICRISQESGILDPDYST 144
Db 11 lqpetgvcsalpsldldrrgaegpe--adrlraayvqeyarlllyq-----lqgsrlm 64
OY 145 GERPSLLSQSALQINSKEPGRFYPASYHNSQTLALGETTPSOLPARQATARATGQSFQ 204
Db 65 gs--aelgssaesargmpgry-----ghlmgt-----gfssrsgmsgdk 102
OY 205 GTTSRAGHLAGEPPAPPPPPPPREFPAPSLGSARELPAPAPAAAAALYSSSTLPAPPR 264
Db 103 tsftr-----plakpayspawssrsrsavdltsrllssahnggsafigavy 148
OY 265 GSPPLAAROGSPTRLQRGSGAPBEGATYAARPGSKPQSRILAKSVSTSPINIVSSA 324
Db 149 ggtcplpmptrpvsfherggaasradydtslpslltggpddtys-----vse 200
OY 325 GLSPIRVTSPPTVOSTISSPDIHQLSSTIGTYATLSPTKRLVNAASEOYSKHSELYATAT 384
Db 201 qlqp-----aaastlyraya-----ykrqass----- 221
OY 385 LQPGSLAAGSRASYSSQGHGLCPRLALQSPENHIDPIYEDRYQKPPAKSLSSQSGDP 444
Db 222 ----gssrag-----gldeapeategr--psrlirapamtlrlqrfqsh 258
OY 445 LPPAHTGTYR---TSTAPSPGVDSVPLQRTGSGHGGQNAAAATFGQASVYAAGPASYAD 501
Db 259 rsrgtgvsvsgaglleparapvrsllsladsgn----- 293
OY 502 PYROLQYCPVSESPYSKSGPALPEGTLARSPIDSIQKDPREFGMRDPELPEYIOMLQH 561
Db 294 -----lpdvrgldsyg-----hrtlqrlss-----gfddldlpsavkylma 330
OY 562 QPSPVOSNAAYLOHLCEGDNKIKAEIRROGGIOLVLDLDRHMTVEHRSACGALRNLY 621

```



```

Db 331 sdprlqylgaaylqhrncysdaaakkgarslqavprlvklfnhangevqrnatgmrntly 390
Qy 622 GKANDDKIALKNCGGIPALVRLRKRTDTLEIRELYTCVLMWSSCDALKMPITODALAV 681
Db 391 d-rvudkhalveengqifellrflrteqdd-elrkvntgllwnssdhkldrlardlq 447
Qy 682 LTNVITIPHSQEMSPLODDRKIOLHSS--QVLRNATGCLRNVSSAGEARRRMECDGL 739
Db 448 ltdlviaprlsppgprl-----lqnaaseelfynatgflrnlsasqatqymtechgyl 502
Qy 740 TDALLVYIQSLGSSSEIDSKTVENCVCILRNLSYRL-----AAETSGOGHMTDELGD 792
Db 503 vdaivtylnaldygcckedsvenavcvtlrnlsyrllydempsalqrlegrr-grdmag 560
Qy 793 LLGGEANGKDAESESQCGKKKKKKKQDDQMDVGPRLPDCAEKPQIGIOMHPSTYKPLRT 852
Db 561 appg-----emwgcftpsrrlrlrelplrtadalftaevskkbpkdlwspqaylgnr 613
Qy 853 LLSBGS-NPDTLEGACAGALONTLAAGSMKWS-VYTRAAYRKEKGLRILVELLRIDNDRYVC 910
Db 614 llrqcelnrnhteaagaqnlftagdiravaylstrlaleqerllnplldrvrtadnqlr 673
Qy 911 AVATALRNALMDVPRKRLICKYAMRDLVNRPRCGNNSNNTAKASMSDQTVTAVCTLEHV 970
Db 674 sltqlllnlrnaankkemsctkvshlklelpr-----svgecprpaevlvalvlnl 728
Qy 971 ITKMMENAKALRDAGALETEKLVGTSKSGDKHSPKVVYKAAQVLNSWMOYRDLRSYKKDG 1030
Db 729 vvaspaardllydgjrlrklvllkktrdpsdkssraasllanlwgsyklhrdfrakg 788
Qy 1031 WSOYHNFV 1037
Db 789 yrkedfl 795

```

RESULT 13
AAB36463
ID AAB36463 standard; Protein; 726 AA.
XX
AC AAB36463;
XX
DT 01-MAR-2001 (first entry)
XX
DE Human plakophilin ppla (PKP1a) protein SEQ ID NO:3.
XX
KW Plakophilin-3; PKP3; catenin-like protein; characterisation; diagnosis;
KW desmosome; epithelial cell; skin disease; dermatological; gene therapy;
KW vaccine.
XX
OS Homo sapiens.
XX
PN WO200066619-A2.
XX
PD 09-NOV-2000.
XX
PF 28-APR-2000; 2000WO-EP04389.
XX
PR 30-APR-1999; 99EP-0870093.
XX
PA (VLAAs) VLAAms INTERUNIVERSITAIR INST BIOTECHNOG.
XX
PI Van Roy F, Bonne S;
XX
DR WPT; 2000-687529/67.
XX
XX
XX Nucleic acids encoding Plakophilin-3 polypeptides, useful for treating
PT skin diseases and disorders of epithelial tissue associated with
PT inappropriate Plakophilin-3 expression and activity -
XX
PS Example 1; Fig 4; 132pp; English.
XX
CC The present invention describes an isolated or recombinant nucleic acid
molecule (I) encoding a Plakophilin-3 (PKP3), from humans, mice and

CC Xenopus laevis. (I) has dermatological activity, and can be used in
 CC gene therapy and for vaccines. (I) and the protein it encodes may be
 CC used in the prevention, treatment and diagnosis of diseases associated
 CC with inappropriate PKP3 expression, such as skin diseases and disorders
 CC affecting epithelial tissue. For example, (I) and vectors containing
 CC (I) and the PKP3 polypeptide may be used to treat disorders associated
 CC with decreased PKP3 expression by rectifying mutations or deletions in
 CC a patient's genome that affect the activity of PKP3 by expressing
 CC inactive proteins or to supplement the patients own production of PKP3
 CC polypeptides. Additionally, (I) may be used to produce PKP3, according
 CC to standard recombinant DNA methodology, by inserting the nucleic acids
 CC into a host cell and culturing the cell to express the protein. (I) and
 CC complementary sequences may also be used as DNA probes in diagnostic
 CC assays to detect and quantitate the presence of similar nucleic acid
 CC sequences in samples, and hence which patients may be in need of
 CC restorative therapy. The PKP3 polypeptides may be used as antigens in
 CC the production of antibodies against PKP3 and in assays to identify
 CC modulators (agonists and antagonists) of PKP3 expression and activity.
 CC The anti-PKP3 antibodies and PKP3 antagonists may also be used to down
 CC regulate PKP3 expression and activity. The anti-PKP3 antibodies may also
 CC be used as diagnostic agents for detecting the presence of PKP3
 CC polypeptides in samples (e.g. by enzyme linked immunosorbant assay
 CC (ELISA)). PKP3 is a catenin-like protein, which is present in desmosomes
 CC and nuclei of epithelial cells. The present sequence represents human
 CC PKP3a (pkp3a), which is used in the exemplification of the present
 CC invention.

Query Match	11.38;	Score 719.5;	DB 21;	Length 726;
Best Local Similarity	29.98;	Pred. No. 6.6e-43;		
Matches 219;	Conservative 108;	Mismatches 271;	Indels 135;	Gaps 23;

```

QY 365 LVHASEYUŠKSHOELIATATLOQRECSLAQ-----SRASYSQ-----HEHLPEL 410
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 46 mmltvkrqšksqssq--sšlsšnrsqmydgldhnygltsršsyuškfgagngswg-- 1000
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 411 RALQŠREHNIPRYEDRYUOKRPYKMSLSQSGDBLPRAHNTGTUTSTAPASPGVDSVQL 4707
      ||| : : : | : : : : : : : : : : : : : : : : : : : : : : : :
Db 101 -----yplnygltkierpdmrfs-----šysqm 123š
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 471 RTGŠHGPRONMAAAT-----FORASYAGPASNAYDPYROLQY-----CPŠVEŠP 515
      | : | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 124 enwšrhyrgsentsgagšdicfmqkškasrsepdylcydprgltkgtlsgkqktcqr 183
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 516 YS-----KSGRALPREGILARPSL--DŠIOKPREEGW-----RPE- 551
      ||| : : : | : : : : : : : : : : : : : : : : : : : : : : : :
Db 184 yšfšycsqgkalkkcprrpscasqgdprvylpšcšk-dšfšhnarasekšcediec 242
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 552 ----PPEYIOMLOHPEPVSOMAAAYLOHLEFGNKIAEIRPQGGJOLVLDLHDHYTE 607
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 243 sgltlrpkavqylsgdekyygatgaylyhntcfqdesakqvyqylgglctklvdlšrpnq 302
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 608 VHRŠAGALRYLVYGRKANDDKRIALKNGGIRALVRLRKTTDEIRELYTVGLVNLSSC 667
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 303 vqgaaagšlmlvf--štšnklettrngjlršvšlltrrgnaešqktglgtlmlnšst 360
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 668 DALKRPITQDALAVLTNAVIIPIHSGWEMŠPŁODDRKLOHŠQVLRNATGCLRNVSSAGE 727
      : : : : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 361 delkeelladprvadlvadlvflpšfcwgqnsnmšrev--dpevfinašgšlmlna-d 417
      : : : : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 728 EARRRMRECDGDTLTLVYIOSALGSSŠIDSŠKYENCŠCILRNTŠYRLAETŠOGQHMGT 787
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 418 agrqlmtrlysgšldšlmlyvqncvaasrčcdkšvencmšcvlnhšylšldaevršryq-- 475
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 788 DELDGLLGCŠEANGKDA----EŠSGCWŠKKKKKKKŠQDQMDŠVGBLPDCAŠRPKŠIOMLH 843
      : : : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 476 -----šeynaršayekšstšgšfnšškdšmmn-šnyd--črlpeetlrpkšgšvlyh 524
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 844 PŠIVAPRYLTŁSEGNPŁŁSAGALONŁLAŠGŠWŠKŠVYIRAAV--RKEGŠPLVŠELLR 902
      : : : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 525 šqšatšrlšmlngškskškdšaleacšqaršlaskyšlšmšgšqšlgšlkeyšlšplaršlg 584
      : : : : | : | : | : | : | : | : | : | : | : | : | : | : |

```


Fri Jul 20 10:09:39 2001

us-09-501-171a-4.rag

Page 16

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2001, 16:00:50 ; Search time 15.64 Seconds
(without alignments)
1577.825 Million cell updates/sec

Title: US-09-501-171a-4

Perfect score: 6377
Sequence: 1 MFARKPPGAALGAMPVPDQ.....PYSELNVTSHVPASPSDWV 1225

Scoring table:
BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 193259 seqs, 20144635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database:

1: Issued_Patents_AA:*
2: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	256.5	4.0	1185	4	US-09-041-886-23
2	194.5	3.1	2842	1	US-07-741-940-7
3	194.5	3.1	2842	1	US-08-289-548A-7
4	194.5	3.1	2842	1	US-08-452-654-7
5	194	3.0	2843	1	US-08-452-655B-2
6	194	3.0	2843	1	US-08-452-655B-7
7	194	3.0	2843	4	US-08-450-582-2
8	194	3.0	2843	4	US-08-450-582-7
9	194	3.0	2843	2	US-08-821-355A-7
10	194	3.0	2843	2	US-08-821-355A-7
11	194	3.0	2843	4	US-09-003-687A-7
12	192	3.0	2843	4	US-07-136-605-7
13	192	3.0	2843	1	US-07-741-940-2
14	192	3.0	2843	1	US-08-289-548A-2
15	192	3.0	2843	1	US-08-452-654-2
16	191	3.0	1068	1	US-08-370-235A-2
17	191	3.0	1068	1	US-08-396-479B-12
18	183	2.9	1004	4	US-08-818-823-12
19	180	2.8	1481	2	US-08-916-352-2
20	180	2.8	1481	2	US-08-616-844-40
21	180	2.8	1481	2	US-08-599-654-40
22	180	2.8	1481	3	US-08-944-868A-40
23	180	2.8	1481	3	US-08-944-423A-40
24	180	2.8	1481	4	US-08-944-496-40
25	174.5	2.7	1075	5	PCR-US94-07297-41
26	173	2.7	984	4	US-08-935-450-2
27	171	2.7	1142	2	US-08-764-870-15
					US-08-993-118-7

28	171	2.7	1142	3	US-08-845-528C-7	Sequence 7, Appl
29	169	2.7	509	4	US-08-860-635A-21	Sequence 21, Appl
30	168.5	2.6	708	1	US-08-396-479B-8	Sequence 8, Appl
31	168.5	2.6	708	1	US-08-818-823-8	Sequence 10, Appl
32	168.5	2.6	739	1	US-08-396-479B-10	Sequence 10, Appl
33	168.5	2.6	739	1	US-08-818-823-10	Sequence 10, Appl
34	165.5	2.6	1003	1	US-08-571-758-4	Sequence 4, Appl
35	165.5	2.6	1003	1	US-08-909-984A-4	Sequence 4, Appl
36	165.5	2.6	1003	1	US-08-909-983-1	Sequence 4, Appl
37	159.5	2.5	534	2	US-08-878-563A-1	Sequence 1, Appl
38	158.5	2.5	1321	2	US-08-317-310A-64	Sequence 64, Appl
39	157.5	2.5	1317	3	US-09-083-521-7	Sequence 7, Appl
40	157	2.5	1312	4	US-09-041-886-19	Sequence 19, Appl
41	153.5	2.4	2441	1	US-08-194-468-2	Sequence 2, Appl
42	153.5	2.4	2441	3	US-08-961-739-2	Sequence 2, Appl
43	152.5	2.4	737	1	US-08-188-582-16	Sequence 16, Appl
44	152.5	2.4	737	1	US-08-646-715-16	Sequence 16, Appl
45	152	2.4	507	4	US-08-860-635A-19	Sequence 19, Appl

ALIGNMENTS

RESULT 1
US-09-041-886-23
Sequence 23, Application US/09041886
Patent No. 6235872
GENERAL INFORMATION:
APPLICANT: Bredesen, Dale E.
TITLE OF INVENTION: Propeptidic Peptides, Dependence
TITLE OF INVENTION: Polypeptides and Methods of Use
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: United States
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/041,886
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 2626
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 1185 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-041-886-23

Query Match 4.0%; Score 256.5; DB 4; Length 1185;
Best Local Similarity 20.8%; Pred. No. 4e-10;
Matches 180; Conservative 87; Mismatches 268; Indels 329; Gaps 40;
QY 32 SF-GTNSGDSSETTETTSALIASYKEOLFETRLTRELAEQIVASQLECKLGSETG 90
DB 34 SFGVSTSSDG-KAKKROTKAKRAVEASTPVYKQGRSE-ETSESESETNAPKTK 91

```

OY 91 SMSSMSAAECPQWQOGGOKNIDELITLTGLEIVDSICRISQEGSILDPDQYISGER --- 1474
Db 92 TGOELPRQSPESDLSDLG-----RSINDGSSDPKRIDODNRSTS 1332
OY 148 PSLISQASALQINS-----KEGSPQ-----YPA 1707
Db 133 PSLTSPGSEVENDSDSSGSLQGPANRYHPRLFPSPQRPDPSTPRQEPASERPSVPT 1922
OY 171 SYHSMQ---TLAIGETTS-----QLARGTQARATQOSTSQ---GTTSAAGHIAEP 2187
Db 193 GHAAPEPPTSMFQAPGAPRRPHOLKPGGTGGVLSGSPMPKRGCGGAASSVGGPBNKGK 2522
OY 219 APP-----PPPREPAPSLGSAFHLPDAPPA-----AAAL 2522
Db 253 HPPPTPTISVSSSGASGAPRTKPTPTTVGGG--NLPSAPRPANPHYTNLPPRLRPL 3102
OY 253 YSSSITLP---APPPGG---SPLAAPQ---GSPPTKLORGSAPEGATYAAPRGSSPKSP 3042
Db 311 NNVAASPPOLGAPQLGPHLPSPYAAQGMGGILPQPEKGPILASPRLSPASSAPAP 3707
OY 305 SRLAKSYTSPIT-----NIYVSAAGLSPITVTS-----PPTVOSTISSPITQLSS 3512
Db 371 MFPPTSSSSSSSSAAASSSSSSSSSABFPASQALPRTPHSFPPPTSLSVSNODPKTTP 4302
OY 352 TIGTYATLS-----PTKRLVHAS-----EQ 3711
Db 431 SLPSQAVMSQGP PPPPPYGRLLANSNAHPCPPESTGAQSTAHHPVSTHHNHQOQOQ 4902
OY 372 YKSHQELATATLQRPDS-----LAAGRSAYSSONH---LGBELRALQ---SPEN-- 4182
Db 491 QOQOQOQHNGNSGPPPPGAPFPHPLEGG---SSHHAHYAMSPSLSLRYPGPAHL 5452
OY 419 --HIDPIYEDRYOKPMPKRSLSQ-----SQGD-----PLPPAHT----- 4502
Db 546 PPHSOVSYSQAGPNCPVPSSSSSSSSSTQCSYPCSHPSQGPQAPYPPFPVPTTS 6052
OY 451 -----GTYRSTAPS--SPGVDSYPLORTG 4732
Db 606 SATLSTVIATVAASSPAGYKTASPPCPPPYGKRADSPGAYKTATPBGYKPG--SPSPRTG 6632
OY 474 SOHGPN---AAATFORASTAAGPAPSNYADPYROLQICPSVESPSKSGPA-----LP 5242
Db 664 TPFGYRGTSPPAGPCTFKRGSPYGP-----GLPAPAGSGLPLRP 7052
OY 525 PEGTIARBP--SIDSIQDPREFGCRDEL-----PEYIQLMHQOFPSVOSMAA 5712
Db 706 PPAAPASPEPLSAQIKKEPAE--EYEPSEVPAPSPSPKPVADV-----PSHASQSA 7592
OY 572 AYLOHICGDKTKIKAEIRQCGIOLVLDLHRMTEVHRSACGALRYLYGKANDOKIA 6312
Db 760 RENKILDGFN-----SCA--RSDLTFVPLEGSKLA 7682
OY 632 LKNCGIDALVRLKTTDLIRE 6552
Db 789 KKRAD---LVEKYAREAEORARE 8082
RESULF 2
US-07-741-940-7
; Sequence 7, Application US/07741940
; Patent No. 5352175
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC

```

```

1  TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
2  NUMBER OF SEQUENCES: 94
3  CORRESPONDENCE ADDRESS:
4  ADDRESSEE: Banner, Birch, McKie & Beckett
5  STREET: 1001 G Street, NW
6  CITY: Washington
7  STATE: D.C.
8  COUNTRY: USA
9  ZIP: 20001-4598
10 COMPUTER READABLE FORM:
11 MEDIUM TYPE: Floppy disk
12 COMPUTER: IBM PC compatible
13 OPERATING SYSTEM: PC-DOS/MS-DOS
14 SOFTWARE: Patent In Release #1.0, Version #1.25
15 CURRENT APPLICATION DATA:
16 APPLICATION NUMBER: US/07/741,940
17 FILING DATE: 19920109
18 CLASSIFICATION: 435
19 ATTORNEY/AGENT INFORMATION:
20 NAME: Kagan, Sarah A.
21 REGISTRATION NUMBER: 32,141
22 REFERENCE/DOCKET NUMBER: 1107.035574
23 TELECOMMUNICATION INFORMATION:
24 TELEPHONE: 202-508-9100
25 TELEFAX: 202-508-9299
26 INFORMATION FOR SEQ ID NO: 7:
27 SEQUENCE CHARACTERISTICS:
28 LENGTH: 2842 amino acids
29 TYPE: AMINO ACID
30 STRANDEDNESS: single
31 TOPOLOGY: linear
32 MOLECULE TYPE: protein
33 ORIGINAL SOURCE:
34 ORGANISM: Homo sapiens
35 IMMEDIATE SOURCE:
36 CLONE: APC
37 US-07-741-940-7
38
39 Query Match 3.1%; Score 194.5; DB 1; Length 2842;
40 Best Local Similarity 18.5%; Pred. No. 5.1e-05;
41 Matches 232; Conservative 164; Mismatches 428; Indels 431; Gaps 49;
42
43 QY 23 SASETSSISP-----GLNLSNGDGSETE-TTSAIIASVKEQF-----LQ 61
44 | : : : |
45 | : : : |
46 Db 102 SVSSRSKSGECSPVPMGSPFRGRGFVNGSRSESTGYLEBLEKERSILLADLDKEEKDMYYAQ 161
47 | : : : |
48 | : : : |
49 QY 62 FRLRLRELA-----EROIVASQLE-----RCKLGSSENGSMSSSAREQFOWQSO 107
50 | : : : |
51 | : : : |
52 Db 162 LQNLTKRIDSLTFENFSIQTDWTRQLEYEAQIVYAMBEQQLGTQDMEKRAQRIARIQ 221
53 | : : : |
54 | : : : |
55 QY 108 DGQKDIEDELTTGLLEVDSCIRSLQESGILDPDYSSTGERPSILSOSALQOLNSKEPQSO 167
56 | : : : |
57 | : : : |
58 Db 222 QLEKDI-----LRIRQLQS-----QATEARSS-----QKHEGSHD 255
59 | : : : |
60 | : : : |
61 QY 168 YPASHVSNQTLALGETTSQLPARGTQARATQGSFSQGTTSAGHGLAPGERAPRRPPRRP 227
62 | : : : |
63 | : : : |
64 Db 236 ---ARQNEGGQGVGIN-----MAISGN-GQGSITRMDH----- 285
65 | : : : |
66 | : : : |
67 QY 228 EPPAFLSGSAFHLPAPAPAAAMALYYSSSTLPAPPRGSPPLAAPQGSPTYLQSGSAP 287
68 | : : : |
69 | : : : |
70 Db 286 -----ETASVLSSSSTHSAPRRLTSHLG-----TKVEVYYS-- 316
71 | : : : |
72 | : : : |
73 QY 288 EGATYAAPRGSSPKQSPSLAKSYSTSSPINITVSSAGISPIRVTSPTVOSTISSPIH 347
74 | : : : |
75 | : : : |
76 Db 317 ---LLSMIGTHDKDMGRITLLAMSSSODSCISMROSGLPILLI----- 356
77 | : : : |
78 | : : : |
79 QY 348 QLSSTIGTYATLSPTKRLVHASEQSKSQDELAYATATLQRGSLAAGSRASYSQHG--H 405
80 | : : : |
81 | : : : |
82 Db 337 -----QLHGGNDKS-----VLLNSRQSKAKARRAS-AAAHNIIH 391
83 | : : : |
84 | : : : |
85 QY 406 LGPE-----LRAQSEPHHIDPIYEDRVVYQKPPMRSLSSQGGDPLPAHTGTGYRT 455
86 | : : : |
87 | : : : |

```

```

Db 392 SQPDKRGRREIRVHLHLLQIRAYCETCMEQEAHERGM-----DQDKNMP----- 438
QY 456 STAPSSPVDSVPLOQTGSOHQHPOANAATFORASYAGPASNYADPYRQLOYCESVESP 515
Db 439 -----APVEH-----OICPAV----- 449
QY 516 YSKSGPALPPEGLTAPSPISDSIQKP-REFGMDELPREVIOMLO-----HOE 563
Db 450 -----CVLMKLSFDEEHRHAMELG-----GLOAIALLQVDCEMYGLTNDHS 493
QY 564 PSYOSNAAYLOHLFCGDKIKAEI-RROGGIQLVDLDDHRTVEHRSAGCALNLVYG 622
Db 494 IILRRRAGALNLTGTGDVANKATLCSMGCRALVLAOLKSESDILOQYASVNLNLS- 552
QY 623 KANDKIALKNCGGIPALVRLKRTTDEIRELYTVGLNLS-CDALKMPI--IOLAL 679
Db 553 RADVNSKKTIREVGSYKALMECALEYKREESTLKSYSALNLSAJCTEKKADICAVDAL 612
QY 680 AVLTNAVITPHSGWENSPLODDRKTQLHSSQVLRNATGCLRVSS--AGEEARRMRREC 736
Db 613 AFLVGTLY-----RSQTNLTAIESGGILRVNSSLATNEDHROILREN 658
QY 737 DGLTDLALVYIOSALGSEIDSKTYENCICILNLSYRLAETSGQHMGTELDGLGCG 796
Db 659 NCLQTLQHLKSHSL-----TYSNAGCTLMNLSAR----- 689
QY 797 EANGKADESSGCGKKKKKKKQSDQMDGVPLDCAEPPKGIOMLHPSIVKPYLTLLSE 856
Db 690 --NPKDOEA-----LMDGAVS-MLKNIJH 711
QY 857 CSNPDLLEAGALONLAA-----GSKMSVYIRAAVKKKGPILVELL- 901
Db 712 SKHKMTAMSGAALRLMANNRPKAKYKADANIMSGSLPFLH---VRKOKALEAEALQOH 767
QY 902 -----RIDNDEYVCAVATALRNMALDVANKELIGKYARLDVHRLPGGNNNNTASKAM 955
Db 768 LSEFPENIDN-----LSPKASHRSQNRHKQSLYGDYV-----DNRNDRSDNF 813
QY 956 SDDTYAVACCTHEVYTKKMEAKALRDAGIEKLVGISKSG-----D 999
Db 814 NNGMNTVLPSPYNTVTLVLPSSSSSGSLDSSREKDSLERERGTGLGNYPATEMPTSS 873
QY 1000 KISPKYVKAASOVNLNMMQYRLRLSLYK---KDGMSQYHFAVASSSTIERDRPRSSST 1056
Db 874 KKLQJSTTAQAIKAYMEVSAIHTSQEDRSSGTTTELHCY---TDENNALRRSSAHT 929
QY 1057 PS--ISPVSEPNNSASAPASPREMISLKERKTDYECTGSNATYHAKGEHTSKDAMT 1114
Db 930 HSNTRYFTSENSNRGCSMPYAKLE---YKRSSNDSLNSVSSSDGYGKRGQ---MKPSIE 983
QY 1115 AONTGISTLYRNSYCAPADEDIKHNQVSAQVPQEPSRKDYETTYOPFONSTRNYDE 1169
Db 984 SYSEDESKF-CSYGGYPADLAKIHSANHMD---NDGELDTPLNYSKYSDE 1033

RESULT 3
US-08-289-548A-7
; Sequence 7, Application US/08289548A
; Patent No. 5648212
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YOSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102

```

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Allegretti, LTD
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/289,548A
FILING DATE: 12-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.46943
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9299
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2842 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: APC
US-08-289-548A-7

Query Match 3.1%; Score 194.5; DB 1; Length 2842;
Best Local Similarity 18.5%; Pred. No. 5.1e-05;
Matches 232; Conservative 164; Mismatches 428; Indels 431; Gaps 49;

QY 23 SASKTSLSL-----GLNTSNGDSSETE-TTSAILASVKBDE-----LO 61
Db 102 SVSSRSGECPVMPGSPRRGFEVNGSRESTDYLELEKERSLLADLDKEKREKDMYAA 161
QY 62 FERLTRELEA-----EROIVASQLE-----RCKLSEGTGSMSSSAEFOFQWQSO 107
Db 162 LQNTLRKIDSLTENFSLQDMTRQLFEARQIRVAMEBQLGTCQDMKRAQRIARLQ 221
QY 108 DGOKDIEDLITGLVLDSCIRSLQESGILDPDYSTGERPSLLSQSALQLNKSPPGSPQ 167
Db 222 QIERDI-----LRIRQLQS-----QATEAERSS-----QNHETGSHD 255
QY 168 YPASYHNSQILALGETTPSQLPARGTQARATGOSFSQGTTSRAGHLAGPPAPAPPPPPPP 227
Db 256 ---AERONEGQVGEIN-----MATSGN-GGGSTTRMDH----- 285
QY 228 EPFAPSLGSAFHLPDAPPAAMAAALVYSSSTLPAPRGSGSPLAPOGSPFKLORGSA 287
Db 286 -----ETASVLSSTTHSAPRLTSHLG-----TKENYYS-- 316
QY 288 EGATYAAPRGSSPKQSPSLAKSYSTSPINIVVSSAGLSPIRVTSPPYVOSTISSPPIH 347
Db 317 ---LLSMLGTHDKDMSRTLMLAMSSSDSCISMROSGLPLLI----- 356
QY 348 QLSSTIGTATLSPTRKRLVHASQYSKHSELYATATLQRPGLASGRASYSQHG--H 405
Db 357 -----QLHGNDKS-----VLGNSRSGKEARARAS--AALNIIH 391
QY 406 LGPE-----LRLAQSPENHIDPIEDRYVOKPPMRSLQSOGDPLPPAHTGTYRT 455
Db 392 SQPDKRGRREIRVHLHLLQIRAYCETCMEQEAHERGM-----DQDKNMP----- 438

```



```

QY 456 STAPSPVDSVPIORRTSGHQPNAAATPGRASVAGPASNADDPYRLOXCPSESP 515
Db 439 -----APVEH-----QICPAY---449
QY 516 YSKSGPALPREGTLARSHSIDSIQDP-REFGWRDPPELVYIOMIQ-----HOF 563
Db 450 -----CYLMKLSFDEEHRHANNELG-----GLOALAEHLQYDCMEYGLTNDHYS 493
QY 564 PSVOSNAAYIQLHLCFGONKIKAEI-KRQSGIQLVLDLHRMTIEVRSACGALRLVYG 622
Db 494 ILLRYAEMALTNLTFFGVANKATLCMKCGKRALVAQLSESEDLQOVLASVLRNLISW- 552
QY 623 KANDNKITALKNCGIPALVRLKRTDLEIRELVGLVNLSS-CDALMKPI--IQDAL 679
Db 553 RADVNSKTLREVGSKVRLMECALVEKKESTLSVLSLWMLSHCHENKADICAVGAL 612
QY 680 AVLINAVIIPHSIGWENSLQDOKRIQLHSSOVLNATGCLRNVS---AGEEARRRREC 736
Db 613 AFLVGTLLY-----RSQNTLALIESGGGILRNVSLSLATNEDHROILREN 658
QY 737 DGLFDALYVIOALSGSEIDSKTYENCVCILRNLSYRLAETSGQGHMGDELDGLLG 796
Db 659 NCLQTLLOHLHLSHL-----TIVSNACGTLNMLISAR-----689
QY 797 EANGKDAESSGCGWKKKKKKKSDQMDGVGPLPCAEPPKGIOMLHPSIVKPYLLTLE 856
Db 690 --NPKDQEA-----LMDGAVS-MLKNLIL 711
QY 857 CSNPTELEGAALQNLAA-----GSMKKSVYIRAVNREKGLPLIVELL- 901
Db 712 SKHKMIANGSAALRNLMANRPARYKKNANIMSPSSSLPSLH---VRQKALEALDAQH 767
QY 902 -----RIDNRRVCAVATLARNALDVNRKELIGKYAMRDVLRHLPNGNSNNTASKAM 955
Db 768 LSEITFDNIDN-----LSPKASHRSKQHRKQSLVGYDF-----DTNHDNRDNF 813
QY 956 SDDVTAVCCITLHEVITIKMENAKALRDAGIEKLVGISKSG-----D 999
Db 814 NTGNNTVSPYLNTVLPSSSSSRGSLDSRSSEKDRSLERBRIGLGNYHPATENPGTSS 873
QY 1000 KHSRNVVAAAOVLNSMOYDLSLYK---KDGKSOYHFAVASSSTIERDQORYSSRT 1056
Db 874 KRGLQISTTAQIAKVMEEVSAIHISQEDRSSGSTTEHCV---TERNLRNRSSAANT 929
QY 1057 PS--TSPRVAPNNSAAPASPREMISLKERKTDYECTGSNATYHGAKGERTSRKDPMT 1114
Db 930 HSNNTNFKSEKSNRTCGMPYAKLE---YKRSSNDLSNVSSSDGYGKRGQ---MKPSIE 983
QY 1115 AONTGISLTYRNSYGAPAEIDIKHNOVSAQVPQEPSEKDYETYPQONSTRNDE 1169
Db 984 SYSEDESKF-CSYGOYPADLAHKIHSANHMD---NDGELDTPINYSLEKYSDE 1033

```

RESULT 4
US-08-452-654-7
Sequence 7, Application US/08452654
Patent No. 5691454

GENERAL INFORMATION:

APPLICANT: ALBERTSEN, HANS
APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner, Birch, McKie & Beckett

```

STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452,654
FILING DATE: 25-MAY-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/741,940
FILING DATE: 08-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107,035574
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2842 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: APC
US-08-452-654-7
Query Match 3.1%; Score 194.5; DB 1; Length 2842;
Best local similarity 18.5%; Pred. No. 5,1e-05;
Matches 232; Conservative 164; Mismatches 428; Indels 431; Gaps 49;
QY 23 SASKTSLSLP-----GLNLSNGDSETE-TTSAIILASVKEQ-----LQ 61
Db 102 SVSRSGECSPPVPMGSPRPGVNGSESTGYLELEKERSLLDLKEEKEDMYAAQ 161
QY 62 FERLTRELEA-----ERQIYASOLE-----ECKLSSETGSSMSMSAREOFOMSO 107
Db 162 LQNLTRIDSLTLENSLQTDMTTRQLEYEARQIRYAMEQGLTCODEMKRAQRIARIQ 221
QY 108 DGQKIDEDLTGLELVDSQIRSLQESGILDPQDYSTGERPSILLSQALQINSKEGFSQ 167
Db 222 QIEKDI-----LRIRQLQS-----QATEARRS-----QKKHETGSHD 255
QY 168 YPASYNSNOTLAIGETTPSQLPARGTQARATGOSFSQGTTSAGHLAGEPAPPPPPPPR 227
Db 256 ---AERQNEQOGVEIN-----MATSGN-GQGSTTMDH----- 285
QY 228 EPPAPISGAFHLPDAPPAAMAAALYSSSTLPAPPRGSPLAAPQGGSPTKLORGGSAP 287
Db 286 -----ETASVLSSTSHAPRRLTSLG-----TIVENVYS-- 316
QY 288 EGATYAPRGSPPKOSPRLAKSYSTSPINIVVSSAGLSPIRVTSPPTVOSTISSPPIH 347
Db 317 ---LISMGTGHDKDMSTLTLAMSSQDSQCSIMRQSGCLPLLI----- 356
QY 348 QLSSTIGYATVLSPTKRLVHASQSKHSQELVYATRLQPSLSLAAGSRATSSQHG--H 405
Db 357 -----QLHGNKDS-----VLGNSRSGSKARARAS--AALHNIIH 391
QY 406 LGPE-----LRAIOSPEHHIDPIEDRYOKPPKRSLSQSGQGLPPAHGTGYRT 455
Db 392 SDDDKRGRREIRVHLLEQIRAYCETCWEQDAHEPGM---DQDKNMP----- 438

```

```

QY 456 STAPSSGVDVPLQRTGSGHGFQNAATFORASTAAGPASVADPYRQLOQCPSESP 515
D 439 -----APEH-----QICPAV-----449
QY 516 YKSGPALPPEGLARSPSTDSIOKDP-REFGRDPELVIOMLQ-----HOE 563
D 450 -----CVLMKLSDEEHRHAMNMLG---GLQAIABLLQVDEMYGLTNDHYS 493
QY 564 PSVOSNAAYLOHLCFGDNKIKAEI-RROGGIQLVDLLDHRTVEYRSAGCALRNLVYG 622
D 494 ITRRRTAGMALTYLTFCDDVANKATLCSMKCMALVLAQLKSEEDLQOYIASVLRNLWS- 552
QY 623 KANDNKIALKNCGGIPALVRLIKTTDLIRLVTGVIMNLSS-CDALKMP1-TQDAL 679
D 553 RADVNSKKTREVGSYKALMECALVEKKESTLSKVSALMNLISAHCTENKADICADVAL 612
QY 680 AVLNNVLIHSGWENSPLODDRKIQHSSQVLRNATGCLRNSS---AGEEARRMRREC 736
D 613 AFLVGTLLTY-----RSQTNLAILLESGGILIRNVSLLIATNEDHROILREN 658
QY 737 DGLTDLALVYIOSALSGSEISDSTVENCVILRNLISYRLAETRSQGHMGTDLDGLCG 796
D 659 NCLQTLILHLKSHSL-----TIVSNAGGTLMNLSAR-----689
QY 797 EANGDAESSGCGKKKKKKKKKODQDGVGPLDCAEPKGIOMLHPSTIVKPYTLTSE 856
D 690 --NPKDOEA-----LWDMGAVS-MELNLIH 711
QY 857 CSNPDTELEGAALONLAA-----GSMKWSYVIRAVYKKEGIPILVELL- 901
D 712 SKIKMTAMGSAALRLNLMANRPKYYDANIMSGSLPSIH---VRKQALAEALDACH 767
QY 902 -----RIDNDRYVCAVATALRNALDVRKELIGRYAMRDVLRLLPGGNSNNTASKAM 955
D 768 LSEFEDINDN-----LSPKASHRSKQRMHQSILYDYVF-----DNRHDDNNSDNF 813
QY 956 SDDTVAVACCTLHEVITTKNENAKALPDAGIGLEKLVGISKSG-----D 999
D 814 NTGNMTVLSPLMTVTLPPSSSSSGSLDSRSREKRLERERIGLGNYPATENPGTSS 873
QY 1000 KHSPPKVVKAASOVLNSMOYRDLRSLYK---KQWMSQYHFVYASSSTIERDRQRPYSSST 1056
D 874 KRELQISTRAQAQIAKMEEYSAIHTEQEDRSSGTTLHCY---TDERNALRSSAAT 929
QY 1057 PS--ISPVRSPPNRSASAPASPREMISLKERKTDYECTGSNATYHGAKGHTSRKDAWT 1114
D 930 HSNVTNFTKSENENFTCSMPYAKLE---YKRSNDLSNVSSSDGYKRGQ---MKPSIE 983
QY 1115 AQNTGISTILRNYSYGAPEADIKHNOVSADPVQPEPSRKDIETIYQPNSTRNDE 1169
D 984 SYSEDESKF-CSYGOYPADLAKHKAHSANHMD---NDGEIDTPTNYSIKYSDE 1033

```

```

ADDRESS: Banner & Wilcoff, Ltd.
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452,655B
FILING DATE: 25-MAY-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,548
FILING DATE: 12-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/741,940
FILING DATE: 08-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107,49964
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9299
TELEFAX: 202-508-9100
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-452-655B-2

Query Match      3.0% Score 194; DB 1; Length 2843;
Best Local Similarity 18.5%; Pred. No. 5.6e-05;
Matches 232; Conservative 164; Mismatches 428; Indels 432; Gaps 49;

QY 23 SASSETSLSP-----GLNTSNDGSETE-TTSAILIASYKOE-----LQ 61
D 102 SVSSSGSGSPVPMGSPFRGTVNGSRESTGYLBELERESILLADDKREKEDWYAAQ 161
QY 62 FERLIRELEA-----EROIVASOLE-----RCKLGETSGMSMSASAEPOFOMOS 106
D 162 LQNLTKRIDSPLIENFSLQDMDTRRQLEYEARQIRVAMEQLGTCQDMKRAORRIARI 221
QY 107 QDGQKIDIELTLTGLELVDSICIRSLQESGILDPDYSTERPILLSQSALQLNSKPGSF 166
D 222 QOIERDI-----LRIRLOLS-----QATEAERS-----QNKHETGSH 255
QY 167 QYPASVHNOTLALGETTPSQLPARGTQARATGQSFQGTISRAGHLACGPPAPPPPPP 226
D 256 D---AERONEGQVGEIN-----MATSGN-GQGSTTRMDH-----286
QY 227 REPAPSLGSAFHLPDAPPAALAAALYSSSTLPAPPGSGPLAPOGSGPTKLQRGSA 286
D 287 -----ETASVLSSTSHSAPRLTSHLG-----TKVENYYS- 317
QY 287 PEGATYAAPRGSSPKQSPSLAKSYSTSSPINIVYSSAGLSPIRVTSPPVQSTISSPT 346
D 318 -----LISMGLTHDKDMSFTLLAMSSSDQSCISMROSGCLPLLI-----357
QY 347 HOLSTIGTAVTISPTKRLVHASFOYKHSOELXATATLQRPGLSAGRASYSOQG-- 404
D 358 -----QLLHGNDKDS-----VILGNSRGSKERARARS--AALAHNI 391
QY 405 HLGPE-----LRLAQSPENHIDPIYEDRYQKPPMRSLSQSGDPLPAHTGTGR 454
D 392 HSPDDKRGREIVLHLLDRIATVACETCWEMQCAHPEGM-----DQKXNMP-----459

```

```

OY 455 TSTAASSGVDSPLOFTSGOHGFONAAAATFORASYAAGPASNADPYRLOLYCPSEVES 514
DB 440 -----APVEH-----QICPAV-- 450
OY 515 PYSKSGPALPEPGLTARSIDSIOKDP-REFGNRDEPLEVIOMLQ-----HQ 562
DB 451 -----CYLMKISFDEEHRHAMNELG-----GLQAIABELLQVDEMYGLTMDHY 493
OY 563 FPSVOSNAAYLOHCFEGDNKIKAEI--RROGGIOLLDVLDHRTMEVHRSGAGLRMLVY 621
DB 494 SITLRRYGMALNTLTFEDVANKATLCSKCMRALVLAQLSESDLOQVYASVLRNLWS 553
OY 622 GKANDNKIALKNGGIPALVRLRKTDTLEIRELYTVGLWNLS--CDALKMPI--IODA 678
DB 554 -RADVNSKKTIREVGSVALMECALVEYKKESTLKSIVLSALMNLHAGCHENKADICAVDA 612
OY 679 LAVLTNAVITPHSGWENSPLODDRKIOLHSSQVIRNATGCLRNVSS--AGEEARRRRE 735
DB 613 LAFVGLTLY-----RSQNTLAIIESGGGLRNVSSLIATNEDHROILRE 658
OY 736 CDGLDALLYIOLASLSSEIDSKTVENCVCILNLSVRLAETSGOQHMGTDLDGLLC 795
DB 659 NNCIOTLQHLKSHL-----TIVSNAGCTLMNLSR----- 690
OY 796 GEANGKDAESSGCGWKKKKKKKSDODWDVGLPDCAPPKGIOMLHPISVYKPYLLLS 855
DB 691 ---NPKDOEA-----LMDMGVNS-MLANLI 711
OY 856 ECSPDUTLEGAGALONTAA-----GSMKWSYITRAVAREKGLPILVELL 901
DB 712 HSKHKMIAMGSAALRLNMANRPAKYKDANIMSPSSLPISLH-----YKQKALAEALDAQ 767
OY 902 -----RINDRVYCAVATLARNALDVNRKELLIGKAYAMDVLRHJPCGNSSNNTASKA 954
DB 768 HILSEFDNIDN-----ISPKASHRSKORHKOSLYGDIYF-----DINRHDNNSDN 813
OY 955 MSDDTVAVCCTLHEVITRKNMENAKALRDAGIEKLVGISKSG----- 998
DB 814 FNTGMATVLPSPITNTVLPSSSSSSKSLDSSRSEKDRSLERERIGLGNYPATENPOTS 873
OY 999 DKHSKRVKAAQVINSWQTRDLRLSLK---KQWISOYHVVASSSTIEROROPYSSSR 1055
DB 874 SKRGLOISTTAAQIAKVAEEVSAIHTSOEDRSSSGSTTELCV---DTERNALRRSSVAH 929
OY 1056 TPS--ISPVRSVPNSRASAAPASPREMISLKERKTDECTGSNATYHAKAEHRSRKDAM 1113
DB 930 THSNYNTFKSNSNRKTCGSMYAKLE---YKRSSNDLSNYSVSSDGYCKRQ---MKPSI 983
OY 1114 TAQNTGISTLYRNSYGAPAEEDIKHNOVSAQVPOEPSRKDYETYOFPONSTRNATDE 1169
DB 984 ESYSEDESKF-CSYGOYPADIAHRIHSANHMD---NDGELDTPINYSIKYSDE 1034

```

```

RESULT 6
US-08-452-655B-7
; Sequence 7, Application US/08452655B
; Patent No. 578366
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Wilcoff, Ltd.

```

```

; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,655B
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,548
; FILING DATE: 12-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107,49964
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 7:
; LENGTH: 2843 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ANTI-SENSE: NO
; US-08-452-655B-7

Query Match 3.0%; Score 194; DB 1; Length 2843;
Best Local Similarity 18.5%; Pred. No. 5.6e-05;
Matches 232; Conservative 164; Mismatches 428; Indels 432; Gaps 49;

OY 23 SASEKTSLSLP-----GLNLSNGGSETE-TTSAILASVKEQF-----LQ 61
DB 102 SVSRSRGECSPVPMGSPFRGFGVNSREESTGYLELEKERSILLDLQKEEKEDMYAQ 161
OY 62 FERLTRELEA-----EROIVASOLE-----RCKIGSETGMSMSASAEQOFQMS 106
DB 162 LQNLTRIDSLPTENFSIQDTMRQLEYEARQIRVAMEBOLGTCODEKRAORIRARI 221
OY 107 QDQKDIIEDELTTGLBLVDSCIRSLQESGILDPDQYSTGERPSLLSQSALQUNSKPSGF 166
DB 222 QOIEKDI-----LRIRQLLOS-----QATEAERS--ONKHETGSH 255
OY 167 QYPAHSHNOALAGETTTSQLEPARGTQARAGQSPSQGTSRAGHLAEPAPPPPPPP 226
DB 256 D---AERQNGQCVGEIN-----MATSGN-GQGSTTRMH----- 286
OY 227 REPPAFSLGSAFHLPAPAPAAALAYSSSTLPAPPRGSGPLAAPPQSGPTKIORGSA 286
DB 287 -----ETASVLSSTSHSAPRLTSHLG-----TKVENYVS- 317
OY 287 PEGATYAAPRGSSPKQSPSRKASVYSTSSPINIVYSSAGLSPIRYTSPPTVOSTISSPI 346
DB 318 -----LLSMGTHDKDDMSNTLLAMSSQDSCISMROSCLPLLI----- 357
OY 347 HQLSSTIGYATISPTKRLVHASEQYSKSHQELIYANATIQORGSILAAGRASYSOHH-- 404
DB 358 -----QILHNDKDS-----VLGNRSRGEARARAS-ALAHNIT 391
OY 405 HLEPE-----LRAQSPHHIDPIYEDRVYQKPMRSLSQSGPLPAPHTGYR 454

```

```

Db 392 HSPDDBKRGRRRLVHLLEQLPACETCWEOMEANEPMK-----DODKNMP----- 439
QY 455 TSTAPSSPGVDVPLQRTGSHGPONMAAATFORASVAGPASNVADPYRQLOCPSVES 514
Db 440 -----APVEH-----QICPAV----- 450
QY 515 PYSKGPALPPECTLARSPSIDIQDP-REFGWRDEPELEVIOMIQ-----HQ 562
Db 451 -----CVLMKLSFDEHHRHAMELG-----GQALAEILOVCEMYGLTNDHY 493
QY 563 FPSVOSMAAYLOHLCFEDNKKIKAEI-RRQGTQLLVLDLDHMTVEVHSAGCALINLVY 621
Db 494 STLRIRAGALINLTFGVANAKATCSMKGCRRALVAQLKSESDLOQVLSVLRNLISW 553
QY 622 GRANDNKIALKNCGGIPALVRLRLTDLRELVYGLVNLIS-CDALKMPI--TODA 678
Db 554 -RADVNSKTLREVGSVKALMECALFEVKKESTLKSVALMNLNSAJCTENKADICAVDA 612
QY 679 LAVLTAAVITPHSGWENSPLODDRKTQLHSSOVLRAATGCLRVSS--AGEBARRRMR 735
Db 613 LAELVETLY-----RSQTLTALIESGGILRVSSLIATNEPHROILRE 658
QY 736 CDGLTALLVYISALGSSEIDSKTEWNCVILRNTSYRLAFTSGQMGHTDELGLLC 795
Db 659 NNCLQLLQHLKSHL-----TYSNACGTLMLNSAR----- 690
QY 796 GEANGDAESSGCGWKKKKKKKQSDQWDVGLPLDCAEPPKGIOMLHPSYKPYLLLS 855
Db 691 ---NPKDOEA-----LMDGAVS-MLKNLI 711
QY 856 ECSNPITLBSAGALONIAA-----GSMKSVYTRAARVREKGLPIVELL 901
Db 712 HSKHAKIAMSAAALNLANRPATKKNANIMSPGSSLSPLH---VRKOKALEAELDAQ 767
QY 902 -----RIDNDRVCAVATATLNNMALDVRNKLIGKYAMRDLVHRLPGGNSNNATSKA 954
Db 768 HLETFEDNIDN-----LSPKASHRSKQHRKQSLYGDYF-----DTNKHDDRSN 813
QY 955 MSDDTYAVCCTLHEVITYTKMENAKALRDAGIEKLVGISKSG----- 998
Db 814 FNTGNNTVLSPLYNTVTLVSSSSSSSGSLDSSREKDRSLEREIGLGNVHPATENPGTS 873
QY 999 DKSPVYVAASOVLNMMOYRDLRSLYK---KDGWSQYHFAASSITTRDQRPYSSR 1055
Db 874 SKGGLDITTAQIAKYMEVSAIHTSQEDRSSGSTEHLCHV---TDRNALRRSSAH 929
QY 1056 TPS--ISPVAVSPNNNSASAPASPREMISLKERKTDYECTGSNATYHGAKEHTSKDAM 1113
Db 930 THSNNTNFKSEKSNRTCSMPYAKLE---YKRSNDLSLNSVSSSDGYGKRGQ---MKPSI 983
QY 1114 TAOCTGISTLYRNSYCAPADIKHNOVSAQVPQEPFSRKDYETTYQFQNSTRNDE 1169
Db 984 ESYSEDESKF-CSYGOYPADLAHKIHSANHMD---NDGELDPINYSLYKSYDE 1034

RESULT 7
US-08-450-582-2
; Sequence 2, Application US/08450582
; Patent No. 6114124
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YOSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102

```

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Wilcoff, Ltd.
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450,582
; FILING DATE:
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/452,655
; FILING DATE: 25-MAY-1995
; APPLICATION NUMBER: US 08/289,548
; FILING DATE: 12-AUG-1994
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107,49964
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-450-582-2

Query Match 3.0%; Score 194; DB 4; Length 2843;
Best Local Similarity 18.5%; Pred. No. 5.6e-05;
Matches 232; Conservative 164; Mismatches 428; Indels 432; Gaps 49;

QY 23 SASKTSLSLP-----GLNTSNGDSETE-TTSAILASYKEDE-----LQ 61
Db 102 SVSSRSGBSPVPMGSPFRGFVNGSRSTGYLELEKERSLLADLDKEKEKDWYAAQ 161
QY 62 FERLTRELEA-----EROIVASOLE-----RCKLSGETGSMSSMSAEQFOWOS 106
Db 162 LONLTRKIDSLPTENFSLOQDMTRROLEVEAROIRVAMEBOLOTQCDMEKRAQRIRARI 221
QY 107 QDGOKIDDELTLTGLELVDSCTRSLOESGLDPQDYSTGERPSLLSSALQNLNKKPGSF 166
Db 222 QOIEKDI-----LRIQLLOS-----QATEAERS-----QNHETGSH 255
QY 167 QYPASYSNQTLAGETTPSQLPARGTQARATGQSFQGTTSRAGHLGEPAPPPPPP 226
Db 256 D---AERQNEGQVGGEIN-----MATSGN-GQGSTTRMDH----- 286
QY 227 REPRAPSLGSAFHLPDAPPAANAALYYSSSTLPAPPRGSPPLAAPGSGPTKLQRGGA 286
Db 287 -----ETASVLSSTSTHAPRLTSHLG-----TVEWYYS- 317
QY 287 PEGATYAAPRGSPKQSPRLAKSYSTSPINIVSSAGLSPIRYVTSPTVQSTISSPI 346
Db 318 -----LISMGLTHKDDMSKTLAMSSSDSCISMROSGLPLIT----- 357
QY 347 HOLSTIGTATLSPTRKRVASDOYKSHSQELATATLQRPGLSIAGRASYSSQHG-- 404
Db 358 -----QLHGNDKDS-----VLGSRGSKKEARAAAS--AALHNII 391
QY 405 HLGPE-----LALQSPENHIIDPIYEDRYQKPPMRSLSQSGDPLPRAHGTTR 454

```

```

Db 392 HSQPDKRGREIRVHLHLEIRAVCEWMOAHERGM-----DQKNRP-----439
Qy 455 TSTAESSPGVDSVPLQRTSGOHGPNMAAATFORASTYAAGPASYADPYROLQYCPSESVES 514
Db 440 -----APEH-----OICPAV--450
Qy 515 PYKSGPALPPEGLIARSPSIDSTQKDP-REFGWRDPPELPEVIOMLQ-----HQ 562
Db 451 -----CVMKLSFDEBRHAMNMLG-----GLOAIABLLDYDCEMGLTNDHY 493
Qy 563 PPSVOSNAAYLOHLCFEGDNKIKAEI--RRGGIOQLVDLDHRRATEVHRSACALRLNLY 621
Db 494 SITLRRYGMALTNFTFEDVANKATFCGKCMRALVAQLKSESEDDIQQVIAVLRNLISW 553
Qy 622 GKANDNKIALKNGGIDALVYLLKTTDLIRLVGVNLNLS--CDALKMPI--IQDA 678
Db 554 -RADVNSKTLREVGSVKALMECALVEKKESTLKSVALNLSAHCTENKADICAVDGA 612
Qy 679 LAVLNNAVITPHSGWENSPLODDRIOHSSQVLRNATGCLRNS--AGEEARRRMR 735
Db 613 LAFVGTITY-----RSQTNITLAIIESGGILLRNVSLLIATNEHRQILRE 658
Qy 736 CDGLTDALLVYIOSALGSSEIDSKTVENCVCILRLSYRLAETSGOQHMGCTDELGLLC 795
Db 659 NNCLOTLQHLKHSI-----TIVSNACGTLNMLNLSAR-----690
Qy 796 GEANCKDAESSGCKGKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 855
Db 691 ---NPKDDEA-----LMDGAVS--MLKNLI 711
Qy 856 ECSNDPTLEGAAGALONLAA-----GSMKSVYIRAAVKEKGLPLIVELL 901
Db 712 HSKHKIMAGSAAALRNLMANRPAYKKNANIMSSSLPSLH-----VAKQALAEALDAQ 767
Qy 902 -----RIDNDRVCAVATLRNMLDVRNKEIGIYARMDLVRLHPLGNNSSNNTASKA 954
Db 768 HLETFEDIDN-----LSPRASHRSKORHOSLYGDYVF-----DINRHDNNSDN 813
Qy 955 MSDDTVAVCCILHEVITRNKNEMAKALDAGIEKLVGISKSG-----998
Db 814 FNTGMVTLSPYLTNTVLPSSSSSGSLDSRSEKDRSLERKIGLGNYPATENEGTS 873
Qy 999 DKHSRKYVKAASQVLNSMWOYRDLNLSYK---KDGMSQYHFVASSSTIENDRORPYSSSR 1055
Db 874 SKRGIOISTTAQIAKIVAEYSALHTSQEDSSSGSTTELHCY---TDERNALRRSSAH 929
Qy 1056 TPS--ISYVRSVPPNRSASAPSPREMI SLKERTDYECTGSMATYHAKGEHRSKRDAM 1113
Db 930 THSNVTYNTKSENSNRTOCMPYAKLE---YKSSNDLSINSVSSSDGYKRGQ---MKPSI 983
Qy 1114 TQONTGISITLVNSYGAAEDIKHNOVASOAPPOPEPSKDYETQOPFONSTRNRYDE 1169
Db 984 EYSSEDESKF-CSTGYQYPAIDIAKHISANHMD---NDGELDTPIVYSIKYSDE 1034

```

RESULT 8
US-08-450-582-7

Sequence 7, Application US/08450582

Patent No. 6114124

GENERAL INFORMATION:

APPLICANT: ALBERTSEN, HANS

APPLICANT: ANAND, RAKESH

APPLICANT: CARLSON, MARY

APPLICANT: GRODEN, JOANNA

APPLICANT: HEDGE, PHILIP J.

APPLICANT: KINZLER, KENNETH

APPLICANT: KINZLER, GEORGE

APPLICANT: MARKHAM, ALEXANDER F.

APPLICANT: NAKAMURA, YUSUKE

APPLICANT: THLIVERIS, ANDREW

TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC

TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS

```

NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESS: Banner & Witcoff, Ltd.
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/450,582
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/452,655
FILING DATE: 25-MAY-1995
APPLICATION NUMBER: US 08/289,548
FILING DATE: 12-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/741,940
FILING DATE: 08-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.49964
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
US-08-450-582-7

```

Query Match 3.0%; Score 194; DB 4; Length 2843;

Best Local Similarity 18.5%; Pred. No. 5,6e-05; Matches 232; Conservative 164; Mismatches 428; Indels 432; Gaps 49;

```

Qy 23 SASEKTSLSLP-----GLTNSNGDSERE-TTSMILASVYKQF-----LQ 61
Db 102 SVSSRSGECSVPVDMGSPRRGFVNGSRESTDGYLELEKERSLLADLDKEEKDWTAAQ 161
Qy 62 FERUTRELEA-----EROLVASOLE-----RCKLSEFTGSMSSSAEFOFOWS 106
Db 162 LQNTKRIIDPLITENFSLQTDMTFRQLLEYEAQIRVAMEEQGLTQDMEKRAQRIARI 221
Qy 107 QDGQKDIIEDELTGTELVNDCIRSLQESGILDPQDYSTGERPSLLSQALQNSPREGS 166
Db 222 QQLEKDI-----LRIQLLOS-----QATEAERS-----ONKHETGSH 255
Qy 167 QYASVHSNOTLALGETTSPQLPARGTQARNGQSPSQSGTTRAGHLAPPEAPPPPPP 226
Db 256 D---AERUNGQGVGEIN-----MAVSGN-GQGSTTRMDH-----286
Qy 227 REPPAPSLGSAFLPAPPAAMAAALYSSSTLPAPPRGSPPLAAPQGSPTKLRGSGA 286
Db 287 -----EFASVLSSTSTHAPRLRLSHG-----TKVEWYS-317
Qy 287 PEGATYAAPRGSSPKQSPSLAKSYSTSPINIVSSAGLSPIRYTSPPTVOSTISSSPT 346
Db 318 -----LLSMGTGHDKDDMSRTLLAMSSODSCISMROSCLPLLI-----357
Qy 347 HQLSSTIGTATLSPTRKLVHASEQYSKHSQELXATATLQROGSLAAGRASYSQHG--404

```


QY 515 PYSKSGPALPEPTLARSPSIDISKDP-REFGWRDELPETVIOMLQ-----HQ 562
 Db 451 -----CVMKLSFDEBHRHANNELG-----GLQAIALLQVCEMYGLTNDHY 493
 QY 563 FFSVQSNAAVYLQHLCEFGDNKIKAEI-RRGGIQLLVYLDLDRHMTVEVHRSACALRNLY 621
 Db 494 SITLRRYAGMALNTLTFEGVANKATLCSMKGCMRALVAQLKSESEDLQOVYASVLRNLSW 553
 QY 622 GRANDDKIALKNCIGIPALVRLRKTDTLEIRELYGVLMNLS-CDALKMPI--IODA 678
 Db 554 -RADVNSKTLREVGSVKALMECALEVKESTLSVLSALMNLASHCTEKAKIDCAVDA 612
 QY 679 LAVLTAIVIIPHSGWENSPLODDRKTQLHSSOVLARNATGCLRNVS--AGEBARRRMR 735
 Db 613 LALVGLTLY-----RSQTNLAIIESGGILLRNVSILATNEDHQILRE 658
 QY 736 CDGLTDLALVYIOSALGSSEIDSKTYENCVCILRNLSYRLAETSGQGHMGTELDGLLC 795
 Db 659 NNCLOTLOHLKSHSL-----TIVSNACGTLNLSAR----- 690
 QY 796 GEANGKDAESSGCGMKKKKKKSSQDQMDVGPLPDCAEPKGIOMLMHPSIVKPYLTLLS 855
 Db 691 ---NPKQEA-----LMDGAVS-MLKNLI 711
 QY 856 ECSNPDTEGAAGALQNLAA-----GSKWSEVYIRAAVREKGLPIVELL 901
 Db 712 HSKRMKIAMGSAALRNLMANRPAYKADANIMSPGSLPSLH---VRKQKALEELDAQ 767
 QY 902 -----RIDNDRYVCAVATALRNALDVNRKELIKYAMRDVLRHLLPGGNSNNTASKA 954
 Db 768 HLESTFNDID-----LSPKASHRSQRKHQSLYGDYV-----DTNRHDNRSDN 813
 QY 955 MSDDTVAVOCCTLEHVTIKNMENKALRDAGIEKLKLGISKSG----- 988
 Db 814 FNTGMVTLVSLYTLVLPSSSSRGSLDSRSEKSLERERIGLCNHPATENPCTS 873
 QY 999 DKHSPKYKKAQVYLNMMQYRDLRLSYK--KDWMSQYHEFVASSSTIERDRQRPYSR 1055
 Db 874 SKRGQISTTAQIAKAMEVSAIHTSOEDSSSGSTELHCY---IDERNALRRSSAH 929
 QY 1056 TPS--ISFVRSVNNRSASAPASPREMISLKERKTDYECTGSNATYIGANGEHSTSRDAM 1113
 Db 930 THSWTYNNTKSENSRPTCSMPYAKLE--YKRSSNDLSINSVSSDYGKRGQ--MKPSI 983
 QY 1114 TAQNTGISTLNRNSYGAAEDIKHNQVSAQVPEPSRKQDETQOPQNSRNRYDE 1169
 Db 984 ESYSEDESKE-CSTGQYPADLAHRIHSANHMD-----NDGELDTPIYNSLKTSD 1034

RESULT 11
 US-09-136-605-7
 Sequence 7, Application US/09136605A
 Patent No. 6140052
 GENERAL INFORMATION:
 APPLICANT: He, Tong-Chuan
 APPLICANT: Kinsler, Kenneth
 APPLICANT: Vogelstein, Bert
 TITLE OF INVENTION: Beta Catenin, TCF-4, and APC Interact to
 FILE REFERENCE: 1107.75741
 CURRENT APPLICATION NUMBER: US/09/136.605A
 EARLIER FILING DATE: 1998-08-20
 EARLIER APPLICATION NUMBER: 08/821,355
 EARLIER FILING DATE: 1997-03-20
 EARLIER APPLICATION NUMBER: 09/003,687
 NUMBER OF SEQ ID NOS: 28
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 7
 LENGTH: 2973
 TYPE: PRT
 ORGANISM: Homo sapiens

US-09-136-605-7

Query Match 3.0%; Score 194; DB 4; Length 2973;
 Best Local Similarity 18.5%; Pred. No. 6e-05;
 Matches 232; Conservative 164; Mismatches 428; Indels 432; Gaps 49;

QY 23 SASKETSLSIP-----GLNLSNGDSSTE-TTSAIIASVGEQ-----LQ 61
 Db 102 SVSSSGECPVPMGSPFRGCVNRSSTGYLEIELEERSILLADLDEKEREDWYAAQ 161
 QY 62 FRLTRLELA-----EROIVASOLE-----RCKLSGETSGSSMSAEDQFQMS 106
 Db 162 LONLKRIDSLEPLTENFSLQDTMTRRQLEYEARQIRVAMEEDLGTQDMERRAQRARI 221
 QY 107 OQOKRIDELETTGLEIVDSCIRSLQESGILDPQDSTGERPSLLSOSALQNSKEPSF 166
 Db 222 QQLERDI-----LRIRQLQS-----QATERNS-----QNKHEIGSH 255
 QY 167 QYPASYHNOQLALGETTPRSQPLPARGQARATGQSFQGTTSRAGHLAGEPAPPPPPP 226
 Db 256 D--AERQEGGQVGEIN-----MATSQN-GGGSTTRMDH----- 286
 QY 227 REPFPASLSAHLHPDAPPAALAAALYYSSSTLPAPPRGSPILAPGGSPTKXIQRGSSA 286
 Db 287 -----ETASVLSSTHSAAPRLTSLG-----TKEMVYS- 317
 QY 287 PEGATVAPRGSSPKOSPRLAKSYSTSPINIVSSAGLSPIRYTSPPTVOSTISSPI 346
 Db 318 -----LISMGLHDKDKMSRTLLAMSSSDSCISMRQSCPLLI----- 357
 QY 347 HOLSTIGTATLSPYKRLVHASSEYKSHQELVATLQRPGLSAGSRASYSSQHG-- 404
 Db 358 -----QLHGNKDS-----VLGNSRSGKEARARAS--AALHNI 391
 QY 405 HLGPE-----LRAQSPENHIDPYEDRYVQKPPMRSLSSQDDPLPAPATGYR 454
 Db 392 HSQPDKKRRREIRYVHLLEQIRACETCEWQEAHEPEM--DQKNPMA----- 439
 QY 455 TSTAPSPGVDSVPLQRTSGHQPQMAAATFORASAYAGPASNYADPYQLQYCPSEVS 514
 Db 440 -----APVH-----QICPAV-- 450
 QY 515 PYSKSGPALPEPTLARSPSIDISKDP-REFGWRDELPETVIOMLQ-----HQ 562
 Db 451 -----CVMKLSFDEBHRHANNELG-----GLQAIALLQVCEMYGLTNDHY 493
 QY 563 FFSVQSNAAVYLQHLCEFGDNKIKAEI-RRGGIQLLVYLDLDRHMTVEVHRSACALRNLY 621
 Db 494 SITLRRYAGMALNTLTFEGVANKATLCSMKGCMRALVAQLKSESEDLQOVYASVLRNLSW 553
 QY 622 GRANDDKIALKNCIGIPALVRLRKTDTLEIRELYGVLMNLS-CDALKMPI--IODA 678
 Db 554 -RADVNSKTLREVGSVKALMECALEVKESTLSVLSALMNLASHCTEKAKIDCAVDA 612
 QY 679 LAVLTAIVIIPHSGWENSPLODDRKTQLHSSOVLARNATGCLRNVS--AGEBARRRMR 735
 Db 613 LALVGLTLY-----RSQTNLAIIESGGILLRNVSILATNEDHQILRE 658
 QY 736 CDGLTDLALVYIOSALGSSEIDSKTYENCVCILRNLSYRLAETSGQGHMGTELDGLLC 795
 Db 659 NNCLOTLOHLKSHSL-----TIVSNACGTLNLSAR----- 690
 QY 796 GEANGKDAESSGCGMKKKKKKSSQDQMDVGPLPDCAEPKGIOMLMHPSIVKPYLTLLS 855
 Db 691 ---NPKQEA-----LMDGAVS-MLKNLI 711
 QY 856 ECSNPDTEGAAGALQNLAA-----GSKWSEVYIRAAVREKGLPIVELL 901
 Db 712 HSKRMKIAMGSAALRNLMANRPAYKADANIMSPGSLPSLH---VRKQKALEELDAQ 767
 QY 902 -----RIDNDRYVCAVATALRNALDVNRKELIKYAMRDVLRHLLPGGNSNNTASKA 954

Db 768 H1SETFDNIDN-----LSPKASHRSKQRHKSQSLYGVF-----DTNRHDNRSDN 813

QY 955 MSDDVTAVCCTLHEVITKNNENAKALRDAGIEKLVGISKSGK----- 998

Db 814 FNTGMMTVLSPLYMTVYLPSSSSSRGSLDSRSRSLERERIGICGNHPATENPOTS 873

QY 999 DKHSFKVKAASQVILNSMWOYRDLSLYK---KDGWSQYHFVASSSTIERDRORPYSSSR 1055

Db 874 SKRGQISTTAQJAKVAKVEYSALHTSOEDRSSGSTTELHCV-----TDERALRRSSAAH 929

QY 1056 TPS--ISPVVSPNNRSASAPASPREMISLKERKTYECTGSNATYHGANGERTSRKAM 1113

Db 930 THSNYNTKTSKNSRNCISMPYAKLE--YKRSNDSILNSVSSSDGYGKRQO--MKPSI 983

QY 1114 TAONTGISTLYRNSGAPAEIDIKHNOVSAQPYPOEPRSKRDYETOPONSTRANDE 1169

Db 984 ESYSEDESKF-CSTGYOPADLAKIHSAHMD-----NGELDTPTNYSIKYSD 1034

RESULT 12

US-07-741-940-2

: Sequence 2, Application US/07741940

: Patent No. 5352775

: GENERAL INFORMATION:

: APPLICANT: ALBERTSEN, HANS

: APPLICANT: ANAND, RAKESH

: APPLICANT: CARLSON, MARY

: APPLICANT: GRODEN, JOANNA

: APPLICANT: HEDGE, PHILIP J.

: APPLICANT: JOSLYN, GEOFF

: APPLICANT: KINZLER, KENNETH

: APPLICANT: MARKHAM, ALEXANDER F.

: APPLICANT: NAKAMURA, YUSUKE

: APPLICANT: THILVERIS, ANDREW

: TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC

: TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS

: NUMBER OF SEQUENCES: 94

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: Banner, Birch, McKie & Beckett

: STREET: 1001 G Street, NW

: CITY: Washington

: STATE: D.C.

: COUNTRY: USA

: ZIP: 20001-4598

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: COMPUTER: IBM PC compatible

: OPERATING SYSTEM: PC-DOS/MS-DOS

: SOFTWARE: Patent Release #1.0, Version #1.25

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/07/741,940

: FILING DATE: 19920109

: CLASSIFICATION: 435

: ATTORNEY/AGENT INFORMATION:

: NAME: Kagan, Sarah A.

: REGISTRATION NUMBER: 32,141

: REFERENCE/DOCKET NUMBER: 1107.035574

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: 202-508-9100

: TELEFAX: 202-508-9299

: INFORMATION FOR SEQ. ID NO: 2:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 2843 amino acids

: TYPE: AMINO ACID

: TOPOLOGY: linear

: MOLECULE TYPE: protein

: US-07-741-940-2

Query Match 3.0%; Score 192; DB 1; Length 2843;

Best Local Similarity 18.5%; Pred. No. 7.8e-05;

Matches 232; Conservative 163; Mismatches 429; Indels 432; Gaps 49;

QY 23 SASKTSLSIP-----GLNLSNGCSETE-TTSAIILASVKEQF-----LQ 61

Db 102 SVSRSRGECSPVPMGSPFRPGFVNSRSTGYLLEBEKERSLLDLDKEEKXMDYQAQ 161

QY 62 FERTITRELEA-----ERQIVASQLE-----RCKLGSFTGSMSSMSAEQFQWOS 106

Db 162 LQNLFTIRIDSLPLETFENFSQTLTRQLEYEARQJRVAMEEQLGTCOMERAKQRIARI 221

QY 107 QDGKDIDELDTTGLVUDSCIRSLQESGILDPDQYSTEGERPSLLSOGALDINSKPEGSF 166

Db 222 QOLEKDI-----LRITROLLOS-----QATBAERS-----QNKHETGSH 255

QY 167 QXPASHVNOTIAGETTPPSQLPARGTQARATGQSFQSGTTSRACHLAGPAPAPPPPPP 226

Db 256 D--AEKONEGQVGEIN-----MATSGN--GGGSTRMDH----- 286

QY 227 REPFAPSLGSAEHLDPAPAAAAALYYSSSTLPAPRPGGSPLAAPGGSPTKLORGSA 286

Db 287 -----ETASVLSSSTSHAPRRITSHLG-----TKVEMVYS- 317

QY 287 PEGATYAAPRGSSPKQSPRLAKYSTSPINIVVSSAGLSPIRTSPYQSTISSSPI 346

Db 318 -----LLSMLGTHDKDMSRLLAMSSQDSCTISKQSGCPLLI----- 357

QY 347 HOLSTIGTYATLSPTKRLVHASEQYKHSQELVATATLQRRGSLAASRASYSOHG-- 404

Db 358 -----QULHGNKDS-----VLGNSKSGREAKARS-AALHNIT 391

QY 405 HLGPE-----LRALQSEHHIDPIYEDRYQKPPMSLSQSGDPLPPAHGTYYR 454

Db 392 HSQPPDKRGREIRVHLHLEQIRAYCETWEQEAHEBGM-----DQDKNMP----- 439

QY 455 TSTAASSPGVDVSPLORTGSOHQPCQMAAATFQRASTYAGASNADRYRLOQCPSEVS 514

Db 440 -----APVEH-----QICPAV-- 450

QY 515 PYKSGPALPBBGLARSPSIDSIQKDP--REFGWRDPELPEYIOMQ-----HQ 562

Db 451 -----CYIMKLSFDEHRHAMNELG-----GLQAIAMELLOYDCMYGLTNDHY 493

QY 563 FFSVQSNAAVYLQHCFCFDNKKAEI--RRQGIQILVLLDHRMTEVHRSAGLARILY 621

Db 494 STILRRYGMALTNTLTFGDVANKATLCMKGCMRALVQLKSESEDLQOYVASYLRINLSW 553

QY 622 GRANDKITALKNGGIPALVRLKTTDLIELVELYCVLNLS--CDALMPT--TQDA 678

Db 554 -RADVNSKTLREVSVALMPCALEVKESTLSVLSALMNLHCHTENKADICAVDGA 612

QY 679 IAVLTNAVLIIPHSWENSPLODDRKIQLHSSQVLRNATGCLRNVSS--AGEEARRRMR 735

Db 613 IAFVLGTLTY-----RSQNTLATIESGGGILRVNYSILATNEHDRQILRE 658

QY 736 CDGLTDALLIYQSLASSEIDSKTYENCVCILRLSLRLAELTISOGHGTDELGLIC 795

Db 659 NNCLOTLLQHLKSHSL-----TIVSNACGTLMNLSR----- 690

QY 796 GEANCKDAESSGCMGKKKKKKSOQOMDGVPLPDCAEPKGTOMLWHPSTIVKYTLIS 855

Db 691 ---NPKDEA-----LMDMGAVS--MLNLI 711

QY 856 ECSNPDTEGAAGALONIAA-----GSMKWSYIIRAAVREKGLPIVELL 901

Db 712 HSKHKMIAMGSAALRINLMANRPAAKYKADANIMSPGSLPH---VRKQALEAELEAD 767

QY 902 -----RIIDDRVYCAVATALRMALDVRKKEILGKYAMMDLVHRLPGGNSNNTATSKA 954

Db 768 H1SETFDNIDN-----LSPKASHRSKQRHKSQSLYGVF-----DTNRHDNRSDN 813

QY 955 MSDDVTAVCCTLHEVITKNNENAKALRDAGIEKLVGISKSGK----- 998

Db 814 FNTGMMTVLSPLYMTVYLPSSSSSRGSLDSRSRSLERERIGICGNHPATENPOTS 873

QY 999 DKHSFKVKAASQVILNSMWOYRDLSLYK---KDGWSQYHFVASSSTIERDRORPYSSSR 1055

Db 874 SKRGLOISTTAQIAKVMEEVSAIHTSQEDRSSGTTTELHCY---TDERNALRRSSAAH 929
 QY 1056 TPS--ISPVKVPNNNSASAPSPREMIISLKERKTDYECTGSNATYHGAKEGHTSKRAM 1113
 Db 930 THSNYNTFTKSENSNMTCSMPYAKLE---YKSSNDLSVSSNDGCGKRGQ---MKPSI 983
 QY 1114 TAONTGISTLYRNSYGAPAEADIKHNOVSAQVPQEPSPKRDYETQFQONSTNYDE 1169
 Db 984 EASSEDESKF-CSTGYQPADLAHKIHSANHMD---NDGELDPINISLKYSD 1034
 RESULT 13
 US-08-289-548A-2
 ; Sequence 2, Application US/08289548A
 ; Patent No. 5648212
 ; GENERAL INFORMATION:
 ; APPLICANT: ALBERTSEN, HANS
 ; APPLICANT: ANAND, RAKESH
 ; APPLICANT: CARLSON, MARY
 ; APPLICANT: GRODEN, JOANNA
 ; APPLICANT: HEDGE, PHILIP J.
 ; APPLICANT: JOSLYN, GEOFF
 ; APPLICANT: KINZLER, KENNETH
 ; APPLICANT: MARKHAM, ALEXANDER F.
 ; APPLICANT: MAKAMURA, YOSUKE
 ; APPLICANT: THLIVERIS, ANDREW
 ; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
 ; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
 ; NUMBER OF SEQUENCES: 102
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Banner & Allegretti, LTD
 ; STREET: 1001 G Street, NW
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20001-4598
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/289,548A
 ; FILING DATE: 12-AUG-1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Kagan, Sarah A.
 ; REGISTRATION NUMBER: 32,141
 ; REFERENCE/DOCKET NUMBER: 1107,46943
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-508-9100
 ; TELEFAX: 202-508-9299
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2843 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-289-548A-2
 Query Match 3.0%; Score 192; DB 1; Length 2843;
 Best local Similarity 18.5%; Pred. No. 7.8e-05;
 Matches 232; Conservative 163; Mismatches 429; Indels 432; Gaps 49;

QY 107 QDGGKIDIEDELTGILELVDSICIRLSGSIIDPODYSTGERPSILSOLQNLSPKPGCF 166
 Db 222 QOIERDI-----LIRIROLLS-----QATEAERS-----QNNHETGSH 255
 QY 167 QYPASYSNQTALAGETTPPSQLPARGIOARATGSGFSQGTTSRAGHLAGEPPAPPPPPP 226
 Db 256 D---AERONEQGVGEIN-----MATSGN--GQGSTTRMDH----- 286
 QY 227 REPAPSLGSAFHLPDAPPAALAAALYSSSTLPAPPGSGPLAAGGSPTKLORGSA 286
 Db 287 -----ETASVLSSTSHAPRLTSHLG-----TVEWYYS 317
 QY 287 PEGATYAAPRGSSPKQSPSRLAKSYSTISSPINIVSSAGSLPIRTVTSPTVQSTISSPTI 346
 Db 318 -----LISMGLTHDKDDMSRTLLAMSSSDSCISMRSGCPLLLI----- 357
 QY 347 HOLSTICTYATLSPTRKLVASFOYSKHSQELYATATLQRPGLSAGRASYSOHC-- 404
 Db 358 -----OLHGNKDS-----VLGNSRGSKEPARARAS--NALHNIT 391
 QY 405 HLGPE-----LRALQSPENHIDPIEDRYOKRPPKMSLSQSGDPLPPAHTGTYR 454
 Db 392 HSQPDCKRGREIRVHLHLEQIRAVCETCWEOMEAHERGM---DQKKNPW----- 439
 QY 455 TSTAPSSPGVDSVPLQRTGSGQHGPONAAATFORASTAAGPASTADPYROLQYCPSEVS 514
 Db 440 -----APVEH-----QICPAV-- 450
 QY 515 PYSKSGPALPEGTIARSPSIDSIQKDE--REFGWRDPELPFVIOMLQ-----HQ 562
 Db 451 -----CYMLMLSTDEEHRHAMNGLG---GLQAIALLQVDCMYGLTNDHY 493
 QY 563 PPSVQNAAYLQHLFCFQDNKIKAEI--RRQGIQILDVLDLHRTMEVHRSAGALRLVLY 621
 Db 494 STLRRYAGMLTNTLPFGDVANKATLCMKCMRVALYQLKSESEDLQOYVASYLRNLISW 553
 QY 622 GKANDNKIALKNGGCTALYRLKTTDLERILEYGVLMNLS--CDALKMP1--IQA 678
 Db 554 -RADVNSKKTLLREGVSVALMECALVEKVESTLSVLSALNNLAHCETENADICAADGA 612
 QY 679 LAVL#NAVYIIPHSWENSPLODDRKIQHSOVLNATGCLRNS--AGEEARRRRE 735
 Db 613 LAFIVGTLTY-----RSQNTIATIISSGGIIRNYSLSLATIEDHROLIRE 658
 QY 736 CDGLTDALLYIQSALGSEISDSTKTVENCVCILRNLSYRLAETSGQGHMGCTDELGLLC 795
 Db 659 NNCLQTLQHLKSHSL-----TIVSNACGTIMNLSAR----- 690
 QY 796 GEANGKDAESSGCKKKKKKKKSDQMDGCVPLPDCAPPRGIQMLHPSLYKPYLTLS 855
 Db 691 ---NPKDQEA-----LMDGAVS--MLKNLI 711
 QY 856 ECSNPDTLEGAAGALONLAA-----GSWKSXYIRAAVKEGELPIVLELL 901
 Db 712 HSKHKMIAMGSAALRLNLMANPRAYKDNANIMSPSSLSPLH---YKQKALAELEDAQ 767
 QY 902 -----RIDNDRVCAVATALRNALDVRNKLIGKAYAMDVLHRLPGNNSNNTASKA 954
 Db 768 HLETFEDNIDN-----LSPKASHRSKORHKQSLYGDYVF-----DINRHDNRSDN 813
 QY 955 MSDDTYVAVCCITLHEVITKNNMENAKALRDAGIEKLVGISKSG----- 998
 Db 814 FVTGNMTVLSPLYLNTTVLPSSSSSSSGSLDSRSKSEKDRSLERKIGLGNYPATENPGTS 873
 QY 999 DKHSPPVKAASQVLSNMWQYRDLRSLYK---KDWMSQYHFAVSSSTIERDRORYSSSR 1055
 Db 874 SKRGLOISTTAQIAKVMEEVSAIHTSQEDRSSGTTTELHCY---TDERNALRRSSAAH 929
 QY 1056 TPS--ISPVKVPNNNSASAPSPREMIISLKERKTDYECTGSNATYHGAKEGHTSKRAM 1113
 Db 930 THSNYNTFTKSENSNMTCSMPYAKLE---YKSSNDLSVSSNDGCGKRGQ---MKPSI 983

QY 1114 TAONTGISTLYRNSYGAPAFEDIKHNOVSAOPVPOEPSRKDYETYPONSTRANDE 1169
 DB 984 ESYSEDESKF-CSYGQYPADLAHKIHSANMMD-----NDGELDTPTINYSIKYSDE 1034

RESULT 14

US-08-452-654-2
 : Sequence 2, Application US/08452654
 : Patent No. 5691454

GENERAL INFORMATION:

APPLICANT: ALBERTSEN, HANS
 APPLICANT: ANAND, RAKESH
 APPLICANT: CARLSON, MARY
 APPLICANT: GRODEN, JOANNA
 APPLICANT: HEDGE, PHILIP J.
 APPLICANT: JOSIYN, GEOFF
 APPLICANT: KINZLER, KENNETH
 APPLICANT: MARKHAM, ALEXANDER F.
 APPLICANT: NAKAMURA, YUSUKE
 APPLICANT: THILVERIS, ANDREW
 TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
 TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
 NUMBER OF SEQUENCES: 94
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Banner, Birch, McKie & Beckett
 STREET: 1001 G Street, NW
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20001-4598

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/452,654
 FILING DATE: 25-MAY-1995
 CLASSIFICATION: 536
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/741,940
 FILING DATE: 08-AUG-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Kagan, Sarah A.
 REGISTRATION NUMBER: 32,141
 REFERENCE/DOCKET NUMBER: 1107.035574
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-508-9100
 TELEFAX: 202-508-9299
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2843 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-452-654-2

Query Match 3.0%; Score 192; DB 1; Length 2843;

Best Local Similarity 18.5%; Pred. No. 7.8e-05; Matches 222; Conservative 163; Mismatches 429; Indels 432; Gaps 49;

QY 23 SASERTSLSP-----GLNTSNGDSETE-TTSALIASVKEOE-----LO 61
 DB 102 SVSSRSGECSPVWMSGPRRGVNGSRESGYDELEKEKSLILADLDKEKKDMYIAQ 161
 QY 62 FERLTRLELE-----ERQIVASOLE-----RCKIGSETGSMSSMSAEQFOWOS 106
 DB 162 LQNLTKRIDSLPTLENSTLOTDLTRQLEYEARQIRVAMEEQGLTCQDMEKRAQRIARI 221
 QY 107 ODGOKDIEDELTLGLBELVDSICISLOESGILDPDQDSTGERPSLLQSALQLOLSKEGSGF 166
 DB 222 QQTEKDI-----LRIROLLQS-----QATEAERSS-----QNKHETGSH 255

QY 167 QYASVHSQNTALGETTTSQLPARGTOARATGSGFSGGTTSRAGHLGPEPAPPPPPP 226
 DB 256 D---AERONEGOGVEIN-----MANSGN-GGGSTRMDH----- 286
 QY 227 REPFAPSLGSAPFLPDAPPAAMAAALYYSSSTLPAPPRGGSPPLAPGGSPKTLORGSA 286
 DB 287 -----ETASVLSSTSHSPRRLTSHLG-----TKVEMYS- 317
 QY 287 PEGATYAAPRGSSPKQSPSLAKSYSTSSPINIYVSSAGLSPRYTSPYQSTISSPI 346
 DB 318 -----LLSMIGTHDKDDMSRTLLAMSSSODSCISMRQSGCLPLI----- 357
 QY 347 HQLSSTIGTVALSPKRLVHASPOYSKHSQBLVATLQRPQSGISAAGRSYSSQHG- 404
 DB 358 -----QLLGNDRKDS-----VLLGNSRGSKEARARAS-AALHNII 391
 QY 405 HLGPE-----LRALQSPENHIDPIYEDRYQKPPMSLSQSGDPLPPAHTGYR 454
 DB 392 HQQPDQKRRRLIRVLHLEQIRATCEICWQEAHEFGM-----DQKNPMP----- 439
 QY 445 TSTAPSSPGVDVPLQRTGSGHQGPQMAAAATFFQASVYAGPASYADPYROLQYCPYES 514
 DB 440 -----APVEH-----QICPAV-- 450
 QY 515 PYSKSGPALPPEGTILARSPIDSIOKDP-RENGWRDPELPEYIQLO-----HQ 562
 DB 451 -----CVLMKLSFDEEHRHAMELNG-----GLQALAEILOVCEMYGLTNDHY 493
 QY 563 PPSVQNAAYLQHLQCFQDNKIKAEI-RRQGGIOLLVLDLHRMEVRSAGALRNLY 621
 DB 494 SITIKRYAGMALTNLTFQDVANKATLCMSKGMRAVLQAKSESIEDLOQVATVLRNLSW 553
 QY 622 GRANDNKITALKNGGIPALVRLKRTTDLTRELVTGVLNLS-CDALMPTI--IQDA 678
 DB 554 -RADVNSKKTREVSQVAMECALVEKKESTKSVLSALNLSHAETENKADICAVGDA 612
 QY 679 LAVLTNAVYIIPHSQWENSPLODRKIQIOLHSQVLENAIGCLRANS--ACEEARRRRE 735
 DB 613 LAFELVGTLY-----RSQNTLALIESGGGILRNSSSLATVEDRRLIRE 658
 QY 736 CDGLTDALLYIOLASGSEIDSKTVENCVCILRNLSRYLAETISQGHMGTDDELGLLC 795
 DB 659 NNCLOTLLQHLKSHSL-----TIVSNACGTIMNLSAR----- 690
 QY 796 GRANGKDAESSCGWKK 855
 DB 691 ---NPKDQEA-----LWDMGAVS-MLNKLI 711
 QY 856 ECSNPDTLEGAGALQNLAA-----GSMKWSYIIRAAVKEKGLPIVLELL 901
 DB 712 HSKHKIMAGSAAALRNLMARPAKRYKRNANIMSPSSLSPLH-----YRKQALAELEDAQ 767
 QY 902 -----RIDNDRVYCAVATALRNALDVRNKEIGKYAMBDLVHRLPGGNSNNTASKA 954
 DB 768 HLSETFDIND-----LSPKASHRSKORHKOSLYGDVYF-----DTNRHDDNNSDN 813
 QY 955 MSDDIVTAVCTTLEHVTIKNNENAKALRDAGCTEKLQISKSG----- 998
 DB 814 FNTGMNTVLSPLYLNTTVLPSSSSSSSGSLDSSRSSEKDRLEREIRIGLGNHPATENEGTS 873
 QY 999 DKHSPPVYKKAASOVLNYSWQYRDLRSLYK--KDGWSQYHVVASSSTLIEDRQPPSSR 1055
 DB 874 SKRGLOITTAQIAKVAHEVSAIHTSGEDSSSGSTTELHCY-----TDENALRRSSAAH 929
 QY 1056 TPS--ISPVVYSPNNRSASAPAPREMIISLKERKTYDECTGSNATYTGANGCHTSRDM 1113
 DB 930 THSNNTYNTKSENSNRTCSMPYAKLE--YKSSNDLSINYSNDYGGKRGQ--MKPSI 983
 QY 1114 TAONTGISTLYRNSYGAPAFEDIKHNOVSAOPVPOEPSRKDYETYPONSTRANDE 1169
 DB 984 ESYSEDESKF-CSYGQYPADLAHKIHSANMMD-----NDGELDTPTINYSIKYSDE 1034

RESULT 15
 US-08-370-235A-2
 ; Sequence 2, Application US/08370235A
 ; Patent No. 5910418
 ; GENERAL INFORMATION:
 ; APPLICANT: VOGELSTEIN, BERT
 ; APPLICANT: KINZLER, KENNETH W.
 ; APPLICANT: HILL, DAVID E.
 ; APPLICANT: JOHNSON, KAREN A.
 ; TITLE OF INVENTION: ANTIBODIES AND ASSAYS FOR DETERMINING
 ; TITLE OF INVENTION: MUTATIONS IN THE APC GENE
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: BANNER & WITCOFF, LTD.
 ; STREET: 1001 G STREET, N.W.
 ; CITY: WASHINGTON
 ; STATE: DC
 ; COUNTRY: US
 ; ZIP: 20001
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/370,235A
 ; FILING DATE: 01-JAN-1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: KAGAN, SARAH A.
 ; REGISTRATION NUMBER: 32,141
 ; REFERENCE/DOCKET NUMBER: 01107,48688
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202 508 9100
 ; TELEFAX: 202 508 9329
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2843 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-370-235A-2

Query Match 3.0%; Score 192; DB 2; Length 2843;
 Best Local Similarity 18.5%; Pred. No. 7.8e-05;
 Matches 232; Conservative 163; Mismatches 429; Indels 432; Gaps 49;

QY 23 SASKSTSLSP-----GLNTSNGDSETE-TTSAILASYKEOE-----LQ 61
 Db 102 SVSRSGECPVPWSPRRGFVNGSRSTGYLELEKERSLLADLDKREKEDWYQAQ 161
 QY 62 FERUTRELEA-----ERQIVASQLE-----RCKLGSETGSSMSAAEQFOWOS 106
 Db 162 LQNTKRIDILPTENFSLQTDLTRQLEEARQIRVAMEEQGLGTQDMKRAQRARI 221
 QY 107 QDQOKDIEDLTLGLVDSICRSIQESGLDPQDSTGERPRLQOSALQLSKEGSP 166
 Db 222 QQLEKDI-----LRIRQLQOS-----QATEERSS-----QNKHETGSH 255
 QY 167 QYPAVSHNQTLALGETTPSPLPARSTQARATGQSFQGTSTPRAGHLAGPEPAPPPPP 226
 Db 256 D---AERONGGQGVGEIN-----MATSGN-GQGSTTMDH----- 286
 QY 227 REPPAFLSGLAFHLPAPPAALALYSSSTLPAPPPGSPPLADPGSPKTLQRGSGA 286
 Db 287 -----ETASVLSSTSTHAPRRLTSHLG-----TKVENWYS- 317
 QY 287 PEGATYAAPRGSSPKOSPSHLAKSYSTSSPINIVVSSAGLSPIRTVTSPTVOSTISSSPI 346
 Db 318 -----LLSMGLGTHDKDDMKSTLLAMSSQDSCISMROSGCLPLLI----- 357

QY 347 HOLSTIGTATLSPTRKIVHASQYKSHSQELVATATLQRPSTLAAGSRASYSOHG-- 404
 Db 358 -----QLHGDNDKDS-----VILGNSRKSARARAS--AALHNII 391
 QY 405 HLGPE-----LRALQSPHHIDPIYEDRVYQKPPMRSLSQGCLPLPAHGTGR 454
 Db 392 HSDPDKRGREIRVHLHLLEQIRAYCETCWEQEAHPCH-----DDKPNMP----- 439
 QY 455 TSTAASSPGVDSPLORTGSHGPPQNAATFORASVYAGPASNADPYROLQYCPYSES 514
 Db 440 -----APVEH-----QICPAV-- 450
 QY 515 PYSKGPALPPEGTILARSPSIDSIQKP-REPGWRDELPEYIOMQ-----HQ 562
 Db 451 -----CVLMKLSFDEEHRHANEIG-----GLQALAEALLQVDCENYGLTNDHY 493
 QY 563 FPSVQSNAAAYLOHLCFCDNKIKAEI-RRQGGIQLLVLDLDRHMEVHRSAGALRNLY 621
 Db 494 SILRRYAGMALNLTFFGDVANKATLCSMKGRALVLAQLKSESEDLQVYVSLRNLSW 553
 QY 622 GKANDDKTALAKNGGIPALVRLRKTDTLEIRELVYVLMNLS-CDALKMPI--IDA 678
 Db 554 -RADVNSKKTIREVGSVYKALMECALEVKESTLSVLSALMNLASCHTENKADICAVDGA 612
 QY 679 LAVLTNAVLIIPHSQWENSPLODOKTOLHSSQVLRNATGCLNVS--AGEARRRRE 735
 Db 613 LAFVGLTY-----RSQVTLAIIESGGILANVSLSLATNEDHQRIARE 658
 QY 736 CDGLTDLALYVIOALGSSEIDSKTVENCVCILRNLSYLAETSOQGMGTDELGLLC 795
 Db 659 NNCLOTQLQHLKSHSL-----TIVSNCGTLMNLSAR----- 690
 QY 796 GEANGDAASSGCGWKKKKKKKSDQDGVGLPDCAEPPKGIOMLMHPSIVKPYLTLS 855
 Db 691 ---NPKDOEA-----LMDGAVS-MLKNLI 711
 QY 856 ECSPDTELEGAALONLAA-----GSMKSVYIRAAVREKCLPLVELL 901
 Db 712 HSKHKMIAGSAAALRNLMANRPAYKIDANINSPSSLSPLSH--VKRQALEAELDAQ 767
 QY 902 -----RIDNDRVYCAVATALRNALDVRNKLIGKYAMRDVHRLPGGNSNNTASKA 954
 Db 768 HLETFDNDN-----LSPKASHRSKQHRKQSLIGDYF-----DTNRHDNRSDN 813
 QY 955 MSDDTYVAVCCTLHVEYITKNMENAKALRDAGIEKLVGISKSG----- 998
 Db 814 FNTGMATVLSPYLNTTVLPSSSSSRGSLDSRSERKDRSLERERIGLGNVHPATENPST 873
 QY 999 DKHSPKVYKAAQOVLSNMWQYRDLRSYK--KQWMSQYHYVASSSTERDRORYSSR 1055
 Db 874 SKRGILQISTTAQOIAKVMEEVSAIHTSOEDRSSGSTTELDHV---TDERNALRRSSAAH 929
 QY 1056 TPS--ISPYRVSPNRRSAPASPPEMISLKERKDYECTGSSNATYHAKGEHSTRKDM 1113
 Db 930 THSNYNYNFTKSNRSTCMRYAKLE---YKRSSNDLSNYSNDGVYKRRQ--MKPST 983
 QY 1114 TAQNTGISTLYRNSYGAPADIKNQVSAOPVPOEPSRKDYETQYOPFONSTRNRYDE 1169
 Db 984 ESYSEDESKF-CSYGOYPADIAHRIHSANHMDD---NDEGLDLPINYSILKYSDE 1034

Search completed: July 19, 2001, 16:03:58
 Job time: 188 sec

Fri Jul 20 10:09:40 2001

us-09-501-171a-4.ra1

Page 16

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2001, 16:03:20 ; Search time 14.73 Seconds

(without alignments)
2848.810 Million cell updates/sec

Title: US-09-501-171A-4

Perfect score: 6377

Sequence: 1 MFARKPPGALGAMPVDDQ.....PYSELNVENTSHYPASPSDWV 1225

Scoring table: BLOSUM62

Searched: 93435 segs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_33:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	ID	Description
1	1423.5	22.3	969	ARVC_MOUSE	P08203 mus musculu
2	1417	22.2	962	ARVC_HUMAN	O00192 homo sapien
3	1375.5	21.6	911	P120_MOUSE	P30999 mus musculu
4	256.5	4.0	1185	DRPL_HUMAN	P54259 homo sapien
5	226.5	3.6	1183	DRPL_RAT	P54258 rattus norv
6	205	3.2	721	YK82_MYCTU	O10690 mycobacteri
7	202	3.2	1487	ICP4_HSVB	P28925 equine herp
8	201	3.2	817	VRP1_YEAST	P37370 saccharomyc
9	200.5	3.1	1902	SMF1_HUMAN	O14497 homo sapien
10	200	3.1	2845	APC_MOUSE	P61315 mus musculu
11	197	3.1	1461	IE18_PRIVF	P11675 pseudorabie
12	195	3.1	1140	YM96_YEAST	O04893 saccharomyc
13	194.5	3.1	2842	APC_RAT	P70478 rattus norv
14	194	3.0	660	YHLL_EBV	P03181 epstein-bar
15	194	3.0	1575	STJ1_HUMAN	O43426 homo sapien
16	194	3.0	2843	APC_HUMAN	P25054 homo sapien
17	193.5	3.0	781	CTNB_RAT	O29082 rattus norv
18	191.5	3.0	781	CTNB_HUMAN	P35222 homo sapien
19	191.5	3.0	1083	T2D3_HUMAN	O00268 homo sapien
20	191.5	3.0	1896	RBPI_DROME	P04052 drosophila
21	190.5	3.0	1744	TENS_CHICK	O04205 gallus gall
22	190	3.0	1487	ICP4_HSVB	P17473 equine herp
23	189.5	3.0	781	CTNB_MOUSE	O02248 mus musculu
24	188.5	3.0	1273	WE1_YEAST	P38968 saccharomyc
25	187	2.9	578	VAC8_YEAST	P39968 saccharomyc
26	185.5	2.9	2517	NCR2_HUMAN	O29618 h nuclear r
27	183.5	2.9	1125	S24C_HUMAN	P35992 homo sapien
28	183.5	2.9	1306	MSB2_YEAST	P32334 saccharomyc
29	183.5	2.9	1429	EXPA_DROME	O07436 drosophila
30	183.5	2.9	3828	TRX_DROVI	O24742 drosophila
31	183	2.9	781	CTNB_XENLA	P26233 xenopus lae
32	182.5	2.9	813	ARM_MUSDO	O02451 musca domes
33	182.5	2.9	1970	RBPI_HUMAN	P24928 homo sapien

ALIGNMENTS

RESULT 1	ARVC_MOUSE	STANDARD	PRT	969 AA.
34	182.5	2.9	1970	1 RBPI_MOUSE
35	182	2.9	1075	1 NEC3_HUMAN
36	181.5	2.8	818	1 CTNB_URECA
37	180	2.8	1443	1 E75C_DROME
38	180	2.8	1647	1 SN24_HUMAN
39	179.5	2.8	1122	1 HD45_HUMAN
40	179.5	2.8	1367	1 AMYH_YEAST
41	178.5	2.8	1365	1 SUZ2_DROME
42	178	2.8	2090	1 N214_HUMAN
43	177.5	2.8	2805	1 MAPA_HUMAN
44	176.5	2.8	1077	1 HLES_DROME
45	176	2.8	701	1 CG1_HUMAN

P08775 mus musculu
O12968 homo sapien
P35224 urechis cau
P13055 drosophila
P51532 homo sapien
O9uq16 homo sapien
P08640 saccharomyc
P25172 drosophila
P35658 homo sapien
P78559 homo sapien
O02308 drosophila
O13495 homo sapien

RESULT 1
ARVC_MOUSE
ID P08203
AC 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ARMADILLO REPEAT PROTEIN DELETED IN VELLO-CARDIO-FACIAL SYNDROME
DE HOMOLOG (FRAGMENT).
GN ARVCF
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Starzynski-Powit A., Kaufmann U., Urbich C., Zupplinger C.;
RT "The armadillo repeat protein ARVCF, a candidate for the velo cardio
RT facial syndrome forms complexes with cadherin."
RL Submitted (JUN-1999) to the EMBL/GenBank/DBD databases.
CC -!- FUNCTION: INVOLVED IN PROTEIN-PROTEIN INTERACTIONS AT ADHERENS
CC JUNCTIONS (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE BETA-CATENIN FAMILY.
CC -!- SIMILARITY: CONTAINS 10 ARM REPEATS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AJ243418; -!- NOT ANNOTATED_CDS.
DR MGD: MGI:109620; Arvcf.
DR PROSITE: PSS0176; ARM_REPEAT; 3.
KW Repeat; Coiled coil.
FT
FT NON_TER 1
FT DOMAIN 1
FT REPEAT 328
FT REPEAT 377
FT REPEAT 419
FT REPEAT 462
FT REPEAT 503
FT REPEAT 521
FT REPEAT 570
FT REPEAT 616
FT REPEAT 634
FT REPEAT 675
FT REPEAT 720
FT REPEAT 721
FT REPEAT 769
FT REPEAT 770
FT REPEAT 814
FT REPEAT 105378 MW; 91130069E484C5B5 CRC64;
SQ
SEQUENCE 969 AA; 105378 MW; 91130069E484C5B5 CRC64;

Query Match 22.3%; Score 1423.5; DB 1; Length 969;
Best Local Similarity 34.4%; Pred. No. 7.4e-61;
Matches 391; Conservative 153; Mismatches 332; Indels 261; Gaps 42;


```

OY 47 TTSAIIASVKEOEIOLERLITRELEAEROIYASOLERCK-----LGSEFGSMSSMSAEQ 101
DB 1 SAAIIIASVKEOEAEFERLITRELEAEROIYASOLERCK-----LGSEFGSMSSMSAEQ 60
OY 102 --FOWSODGOKIDEJELITLGLVDSCLNSLOESGILDPDOSTGEBLLSOSALQIN 159
DB 61 LVLOGSGFSGSASHA-----TMPAPAEVLEET-VTVE 91
OY 160 SKRPSQYRPSYSHNOTLALGETTSPQLPARCQARATGOSFGOTTSAGHL-AGEP 218
DB 92 EDP-GT--PSSHVSITSEDCIT-----KRTEKVT-KYATVTTRVROVPLGPDG 139
OY 219 AP-----PPPPPREPAPSLGSAFHLDPAP-----PAAAAALYSSSTLPA 261
DB 140 LPLDDGPP-----LGS--FADGFLDRHYLLRGCGGPAATLSRYHSS----- 180
OY 262 PPRGSPPLAAOGGSPYKLOKGSAPBAGATYAPR-----GSSPQOS--PSRLAKSYT 313
DB 181 --GG--GFPDGPESRDIPSYGSLSRGLGVRPRPTGLGPGDGCFTLPGR-REAFPM 233
OY 314 SSPINIVVSSAGLSPIRYVTSPTVOSTISSPIHQLSTIGTYATLSPTKRLVHASQYS 373
DB 234 GSE-----SGPPSGRSL--PEHQAEPY-----LDDTRSLAADDEGG 270
OY 374 KHSQELVATATLORPGLAAGSRASYSQHGHLGPELPAIQSPHHIDPIYEDRVYQKP 433
DB 271 PDLEBDYSTATRRPR-----EYGR-GLARAFEDTADGAGLELEER----- 310
OY 434 MRSLSQSGDPLPAHNGTYRTSTAPSSPGVDSVPLQRTGQHOPQAAAAATPQRASYA 493
DB 311 -----PP--PFAAAPLAQ-----PERGSLGSLDR----- 333
OY 494 GPASVADPYPHOLQYCPSESPYKSGPALPEETLARSPEIDSIQDPREFGRDDELP 553
DB 334 -----VWRSPVDSTRREPR--WRDELP 356
OY 554 EVIOMLQHPFSSVOSNAAYLQHLCEGDNKIKAEIRROGGIQLLVLLDRHTEVHSAC 613
DB 357 EVLAWLRHPVPYKANAAYLQHLCEGNEGIRKRRROLRGLPLVALLDHRRAEVRBRAC 416
OY 614 GALRLVYKANDNKIKAKNGGIPALVRLRKTDTLEIYELVGYMNSGCDALCMP 673
DB 417 GALRLVYKANDNKIKAKNGGIPALVRLRKTDTLEIYELVGYMNSGCDALCMP 475
OY 674 IYODALVLTAAVLIIPHSWENSPLODDRIQLHSSQVLRNATGLNRSVSSAGEARRM 733
DB 476 IYDHLQTLTHVYVPHSGWEREPREDSKPRDAEWITVYFKNTSGLNRSVSSAGEARRM 535
OY 734 RECDGLTALLVYIOSALGSSEIDSKTVENCVCILRLSYFLAETSOGOMGTDE--LD 791
DB 536 RECDGLTALLVYIOSALGSSEIDSKTVENCVCILRLSYFLAETSOGOMGTDE--LD 595
OY 792 GLTCEANGKDAESSGCKKKKKKKSSODQMD--GVGPRPDCAEPKGIOMLHPSPYK 848
DB 596 GLTCEANGKDAESSGCKKKKKKKSSODQMD--GVGPRPDCAEPKGIOMLHPSPYK 653
OY 849 PYLLTSCSNPDTEGAGALONLAGSMKWSYIRAAVKEKGLPIVLELLIDNDYR 908
DB 654 PYLLTSCSNPDTEGAGALONLAGSMKWSYIRAAVKEKGLPIVLELLIDNDYR 713
OY 909 VCAVATALRNALDVNRKELIGKIAMRDVLRHLPDGNNSNTA--SKAMSDIYAVCCT 966
DB 714 VCAVATALRNALDVNRKELIGKIAMRDVLRHLPDGNNSNTA--SKAMSDIYAVCCT 770
OY 967 LHEVITTKMENAALRDAGGIEKLVGSKSGDKHSPVYVAAASOVNSMGOYRDLISLY 1026
DB 771 LHEVITTKMENAALRDAGGIEKLVGSKSGDKHSPVYVAAASOVNSMGOYRDLISLY 827
OY 1027 KKDGSQYHFAVSSSTIRDRQRPYSSRTSPISPV--RVSPNNRSASAPASPRE----- 1079
DB 828 QRDGWTKRKF-QSASTAGPCKPTSSGGFDDSTLPLVCKSLDGEKSNRVDIYPMDLGPD 886
OY 1080 --MISLEKRTDYECTGSNATYHGAKGEHTSRKDMATMONTGISTLYRNSYGAPE 1133

```

```

DB 887 GYATVDRRERRT-----LGSDF-----GD-TSEKELRVYGGV-----YCGPLE 926
RESULT 2
AC ARVC_HUMAN STANDARD: PRT: 962 AA.
ID 000192;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ARADILLO REPEAT PROTEIN DELETED IN VELO-CARDIO-FACIAL SYNDROME.
GN ARVCF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A. (SHORT AND LONG ISOFORMS).
RX MEDLINE=97271559; PubMed=9126485;
RA Sirocki H., O'Donnell H., Dasgupta R., Halford S., St Jore B.,
RA Puech A., Parimoo S., Morrow B., Skoultschi A., Weissman S.,
RA Scambler P., Kucherlapati R.;
RT "Identification of a new human catenin gene family member (ARVCF) from
RT the region deleted in velo-cardio-facial syndrome."
RL Genomics 41:75-83(1997).
CC -!- FUNCTION: INVOLVED IN PROTEIN-PROTEIN INTERACTIONS AT ADHERENS
CC JUNCTIONS.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A SHORT FORM AND A LONG FORM
CC (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: FOUND IN ALL THE EXAMINED TISSUES INCLUDING
CC HEART BRAIN, LIVER AND KIDNEY. FOUND AT LOW LEVEL IN LUNG.
CC -!- DISEASE: GENE DELETED IN VELO-CARDIO-FACIAL SYNDROME (VCFS) AND IS
CC HEMIZYGOUS IN ALL VCFS PATIENTS WITH INTERSTITIAL DELETIONS. THIS
CC HEMIZYGOUSITY MAY PLAY A ROLE IN THE ETIOLOGY OF SOME OF THE
CC PHENOTYPES ASSOCIATED WITH VCFS CHARACTERIZED BY A WIDE SPECTRUM
CC PHENOTYPES, INCLUDING CONOTRUNCAL HEART DEFECTS, CLEFT PALATE AND
CC FACIAL DYSMORPHOLOGY.
CC -!- SIMILARITY: BELONGS TO THE BETA-CATENIN FAMILY.
CC -!- SIMILARITY: CONTAINS 10 ARM REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: U51269; AAC51202.1; -.
CC HSSP: Q02248; 2BCT.
CC MIM: 602269; -.
CC MIM: 192430; -.
CC InterPro: IPR000225; -.
CC Pfam: PF00514; Armadillo-seg; 4.
CC ProSite: PS0176; ARM_REPEAT; 3.
CC Repeat: Coiled coil; Alternative splicing.
CC Domain 8 46 COILED COIL (POTENTIAL).
CC Domain 607 623 NUCLEAR LOCALIZATION (POTENTIAL).
CC Domain 608 611 POLY-ARG.
CC REPEAT 348 387 ARM 1.
CC REPEAT 390 429 ARM 2.
CC REPEAT 433 473 ARM 3.
CC REPEAT 468 508 ARM 4.
CC REPEAT 526 565 ARM 5.
CC REPEAT 575 622 ARM 6.
CC REPEAT 646 682 ARM 7.
CC REPEAT 699 738 ARM 8.
CC REPEAT 739 781 ARM 9.
CC REPEAT 782 826 ARM 10.
CC MEDCUNHSAASIIASVKEOEAEFERLITRELEAEROIYASOLERCK-----
CC BRAODGWSGSGMSGGLPLPAMQOVLV -> MPAELR
CC (IN SHORT ISOFORM).
FT

```



```

Db 62 RRHNGRFVGDADLERQKFSIDLK-----NG-----PODHNHLLSTIPRMOEPGCI 108
OY 152 SOSALQJNSKEGSGFYOPASTYHSNQTALGETTPSOLPANGTOAR--ATGQSSQGTSTK 209
Db 109 VEIYTE--EDEBGA-----MSVSVETD-----GTRRRETTVKKVKTWTR 151
OY 210 AGHLAGPEPPAPPPPPREPPAPSLGSAFHLDPAPAAAALAYSSSTLPAPRGGSPL 269
Db 152 T-----VQPPV-----MG-----PDGLPYDASA VSNNTQTL----- 178
OY 270 AAPGSGPTKLORGSAPEGATYAAPRGSSPKOSPSRLAASYSSTSPINIVSAGLSPI 329
Db 179 -----GRDFRKNNGSGP----- 192
OY 330 RVTSPPYQSTISSPIHQLSSTIGTATLSPTRKLVHASEQYKSHQELXATATLORPG 389
Db 193 -----YVQAGTATLPNPFHYPPDGGRHEDGY-----PG 223
OY 390 SLAAGSRASYSQGHLCP-----ELRALQSPENHIDPIYEDRYQKPPMSLSOSOGDPL 445
Db 224 G-----SDNYGSLRVTIRERTRPSMEGRAPSRD--YGGPOPYRVGSSYD-L 272
OY 446 PPAHTGYRTSTAASSPQVDSV-----PLQRTGSSQHPONAAAAATFORASYAAGA 496
Db 273 HRFPPEPYGLDDQSRMGYDLDYGMMSDYGTARTGTP----- 311
OY 497 SNVADPYQLOLYCPSSVSPYKSGPALPPE-----GTLARSPSIDQND-PR 543
Db 312 -----SDPRRLR-----SYEDMIEEVPDDOYWPAPLAHERGSLA--SIDSIRKMP 359
OY 544 EFGWNPDELPIYIOMLOHOPSVOSNAAYLOHLFCGDNKIKAIRQGGTOLVDLDD 603
Db 360 PSNNRQPLPPIVIMLGRDLAVASNAAYLOHLCYNDKDKTVAUKGIPILVGLDH 419
OY 604 RMEYHNSACALRNLYGKANDNKATLAKCGGIPALVRLKRTTLELRELYTVGLVN 663
Db 420 PKKEVHLGACGALKNISGR--DODNKTAIKNCDEVPAVLRKARAMDLEVTGTLNM 478
OY 664 LSSCDALKMPITIDIALVLTNAVITPHSGWNSPLQDDRKTOLHSSOVLNRNATGCLNVS 723
Db 479 LSSHDSIMEIYDHALHATLDEVITPHSGWREPRENECKPRHIMESVLTNTAGCLNVS 538
OY 724 SAGEBARRRMECDGLDALLYTQIALGSSIDSCTYENCVCILRNLYSLAETSQOQ 783
Db 539 SERSPARRKRECDGLVALLFIYQALGQKDSKLEVCCLRNLSYGVHREIQO-- 596
OY 784 HMGDELDELGLGCEANGDASSGCGWKK--KKKKKSDQDDVGPPLDCAEPKKGIOML 841
Db 597 ---ARVQELPVTYANSTGPHRAASCFGAKKGGKPKTEDPANDTVDPFKRTSPARKYEL 653
OY 842 WHPSTIVKPYLTLLSECSNPDTELEGAAGALQMLAAGSMWYVIRAAVKEKGLPILVELL 901
Db 654 FQPEVVRITISLKEKNPALLEASAGAIQNLGGRMYIGYIISALRQEKALABARELL 713
OY 902 RIDNDRVCAVATALRNALDVRNKLIGKYAMDLYHRLPGNNSNNTASKMSDQTVT 961
Db 714 TSEHERVYKASGALRNLAVDARNKELIGKHARPNLYKNLPGGOON--SSWNSSEDTVV 770
OY 962 AVCCGLHEVITKMMENAKALDAGIEKTVGISKSGKHSBPKVYKAAVQSLNSMMOYR 1021
Db 771 SILNTINEVIENLEAKAKLETOGIEKTVLINS--GNRSEKEVRAALVLOQIMNGYKE 828
OY 1022 LRSLYKKGWQYHFVASSSTIERDR--ORPYSSSRTPSISFVSPNNRS 1070
Db 829 LRKPLEKGGWKKSDFOVNLNNAASQSSHSYDSTLPLIDRNQSDNNYS 878

```

```

DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ATROPHIN-1 (DENTATORUBRAL-PALLIDOLYSIAN ATROPHY PROTEIN).
GN DRPLA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cerebellum, and Brain;
RX MEDLINE=95144175; PubMed=7842016;
RA Nagafuchi S., Yanagisawa H., Ohsaki E., Shirayama T., Tadokoro K.,
RA Inoue T., Yamada M.;
RT "Structure and expression of the gene responsible for the triplet
RT repeat disorder, dentatorubral and pallidolysian atrophy (DRPLA).";
RL Nat. Genet. 8:177-182(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=9626214; PubMed=8965642;
RA Margolis R.L., Li S.-H., Young W.S., Wagster M.V., Stine O.C.,
RA Kidwai A.S., Ashworth R.G., Ross C.A.;
RT "DRPLA gene (atrophin-1) sequence and mRNA expression in human
RT brain.";
RL Brain Res. Mol. Brain Res. 36:219-226(1996).
RN [3]
RP SEQUENCE OF 470-725 FROM N.A.
RC TISSUE=Brain cortex;
RX MEDLINE=9315145; PubMed=7825628;
RA Li S.-H., McInnis M.G., Margolis R.L., Antonarakis S.E., Ross C.A.;
RT "Novel triplet repeat containing genes in human brain: cloning,
RT expression, and length polymorphisms.";
RL Genomics 16:572-579(1993).
CC -1- TISSUE SPECIFICITY: THE LEVELS ARE RELATIVELY HIGH IN THE BRAIN,
CC Ovary, Testis and Prostate. Lower levels are detected in the
CC liver, thymus and leukocytes.
CC -1- POLYMORPHISM: THE POLY-GLN REGION OF DRPLA IS HIGHLY POLYMORPHIC
CC (7 TO 23 REPEATS) IN THE NORMAL POPULATION AND IS EXPANDED TO
CC ABOUT 49-75 REPEATS IN DRPLA PATIENTS. LONGER EXPANSIONS RESULT IN
CC EARLIER ONSET AND MORE SEVERE CLINICAL MANIFESTATIONS OF THE
CC DISEASE.
CC -1- DISEASE: DEFECTS IN DRPLA ARE THE CAUSE OF DENTATORUBRAL-
CC PALLIDOLYSIAN ATROPHY, AN AUTOSOMAL DOMINANT NEURODEGENERATIVE
CC DISORDER CHARACTERIZED BY A LOSS OF NEURONS IN THE DENTATE
CC NUCLEUS, RUBROD, GLOBUS PALLIDUS AND LYS/BODY. CLINICAL FEATURES
CC ARE MYOCLONUS EPILEPSY, DEMENTIA, AND CEREBELLAR ATAXIA. ONSET OF
CC THE DISEASE OCCURS USUALLY IN THE SECOND DECADE OF LIFE AND DEATH
CC IN THE FOURTH.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: D31840; BAB0626.1; -.
DR EMBL: U23851; AAB50276.1; -.
DR HSSP: L10377; ? NOT_ANNOTATED_CDS.
DR HSSP: P00651; 1LRA.
DR MIM: 125370; -.
DR InterPro: IPR002951; -.
DR PRINTS: PR01222; ATROPHIN.
KW Triplet repeat expansion; Polymorphism.
KW SER/GID-RICH (MIXED CHANGE).
FT DOMAIN 73 82 POLY-SER.
FT DOMAIN 302 305 POLY-PRO.
FT DOMAIN 376 382 POLY-SER.
FT DOMAIN 386 397 POLY-SER.
FT DOMAIN 442 447 POLY-PRO.
FT DOMAIN 479 483 POLY-HIS.
FT DOMAIN 484 497 POLY-GLN.
FT DOMAIN 504 507 POLY-PRO.
FT DOMAIN 564 574 POLY-SER.

```



```

RESULT 7
ID ICP4_HSVB STANDARD; PRT: 1487 AA.
AC P28925;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DE 01-FEB-1994 (Rel. 28, Last annotation update)
DE TRANS-ACTING TRANSCRIPTIONAL PROTEIN ICP4 (155 KDA IMMEDIATE-EARLY
  PROTEIN).
DE IE OR 64.
OS Equine herpesvirus type 1 (strain Abdp) (EHV-1).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicelloviruses.
OX NCBI_TaxID=31520;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92295566; PubMed=1318606;
RA Telford E.A.R., Watson M.S., McBride K., Davison A.J.;
RT "The DNA sequence of equine herpesvirus-1.";
RL Virology 189:304-316(1992).
CC -1- FUNCTION: THIS IE PROTEIN IS A MULTIFUNCTIONAL PROTEIN CAPABLE
CC OF MIGRATING TO THE NUCLEUS, BINDING TO DNA, TRANS-ACTIVATING
CC OTHER VIRAL GENES, AND AUTOREGULATING ITS OWN SYNTHESIS.
CC -1- SUBCELLULAR LOCATION: NUCLEUS OF INFECTED CELLS.
CC -1- PTM: A LONG STRETCH OF SERINE RESIDUES MAY BE A MAJOR SITE OF
CC PHOSPHORYLATION.
CC -1- SIMILARITY: BELONGS TO THE ICP4/IE140/IE180 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC -----
CC EMBL: M86664; AAB02515.1;
CC EMBL: M86664; AAB02499.1;
CC PIR: A36802; EDBE6.
CC DR HSSP: P04002; 1MFA.
CC DR TRANSFAC: T00401;
CC DR EARLY PROTEIN: Transcription regulation; Trans-acting factor;
CC KW DNA-binding; Phosphorylation; Nuclear protein.
CC FT DOMAIN 181 213 SER-RICH.
CC FT FT 922 931 ARG/LYS-RICH (BASIC).
CC SO SEQUENCE 1487 AA; 154868 MW; 1D4397838D03680D CRC64;

```

```

Query Match 3.2%; Score 202; DB 1; Length 1487;
Best Local Similarity 19.9%; Pred. No. 0.019;
Matches 234; Conservative 134; Mismatches 455; Indels 350; Gaps 50;

```

```

QY 403 -----HGLGPELRALQSPHHIDPIEDRYQ----- 430
DB 265 RSGAASPCAPAPAPASAPSPGGLLPQCARILE-----YLECVREANLAKTLERPEP 320
QY 431 -----KPPMRSLISQSGDPLP-PANTGTYRTSTAPSSGCVDSVPLQRT--GSQHPQ----- 479
DB 321 AGMASPPRSPRHLRPLKQRPKSAALAGASKRRKRRAPR-----IPOTQOAPAEAPQTAW 376
QY 480 -----NAAATFORASAYAGPASNADP-----YRQLOVCPVESPYSGPALTPEPTLAR 531
DB 377 DLDLNNSSQATGAAAAAASAPAAASCAPGVYQRPFLTPSGD--PWQSDP--PPMGVRY 433
QY 532 SPISDITQKDRREFGWRDPELVYIOMLOHOFPSVSNAAAYIOLGFCGDNKTKAETIRQ 591
DB 434 GGTGGS-----RDGLMDDEP-----IVLAASRYAEQAQPVFVPEWEMDSTKYQNALV--- 481
QY 592 GGIQLLDLDLHRTVEYHRSACGALRNIVYKANDDKIALKCGGIPALVRLRTTDL 651
DB 482 -----RMVFPSREMSWLN-----SKLSGODONLQFCQKRTIAPR----- 518
QY 652 EIRELVTVGLWNLSGCDALKMPITIDALAVLTNAVITPHSGWENSPLODDRKIQLHSSQV 711
DB 519 GHGSEFTIGSVAN-----PLPHIGDMAAGNALMALPAA--ASVAMSRRYDTQKSF 569
QY 712 LRNATGCLRNVSSAGEARRRRECDGLTDLALVYIOALGSS-----EIDSKTV 761
DB 570 LQSLRRAYADMAVPRDEAGR-----DSLAAVAGYPAQAAAAAASQOQPEAPASVRYREA 625
QY 762 ENCYCIL-----RNLSTRLAETSGOQHGMDLDELGLGCEANGKDAESSC---WGKKK 814
DB 626 YTRCAALGPRKRAAAAAAGTAPRPSAFRLREL-----GDACVLAQCAVFEALLR 678
QY 815 KKSQODQDGVGP--LPPCAEPKGIOMLHMPSTVKRYTLTLLSPCSNPDLLEGAAGLON 872
DB 679 LRGAASAVPGIDPSEITSPACPEAL-----CSNPAGLETAUSLYE 720
QY 873 LAAGSMKWSYITRAVYKREKGLPTLVELLRIDNDRVCAVATLRNNALDVRNKKELIGY 932
DB 721 LRD-----LVERARLIGD----- 733
QY 933 AMRDLVHRLPGGNNNNNTASKAMSDDTVTAVCTLHEVITTKMBENAKALRDAGIEKLV 991
DB 734 --SDPTHTL--GSDRLRAVRA-----VLVYARTVAVLRYVNEGARARASAMTVQDAVE 784
QY 992 GISKSGDKHSPKYVKAASQVLNMQYRLRLSKKDKSGOYHFAVSSSTIERDRQRY 1051
DB 785 SIPLVGMIGEAV-----SLAPPTRSQO---PSSSS---PGGEPP 820
QY 1052 SSSRT-----PSISPVRYSPNNRSASAPASPREMISLKER-----KTYDECTGSNAT 1098
DB 821 SGAASAEISQIOTLPPLMP--TVPEKQSATVPSHSQSPQHSQSGGAGATTATTCRATQ 878
QY 1099 YHAKGEHTSKDKMTQNTGISTLYRNSTGAPAEADIKHNOVSQAQVYQPSKDIETTY 1158
DB 879 NARSRGQOHQOKARSPQ--AAASPAHLISQAMPSSSDDDRAIHGRPGKSGKRRS--EPLE 936
QY 1159 P-----FQNSTRYNDESFEDQVHHPRP 1182
DB 937 PAAGAGASATSSASAKGTDPSPYDS-----PPA 965

```

```

RESULT 8
ID VRP1_YEAST STANDARD; PRT: 817 AA.
AC P37370; Q06133;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE VERPROLIN.
GN VRP1 OR MDP2 OR END5 OR YLR337W OR I8300.13.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

```

OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RC SEQUENCE FROM N.A.
 RX STRAIN-A364;
 RA MEDLINE=95058201; PubMed=7968536;
 RT Donnelly S.F.H., Pocklington M.J., Pallota D., Orr E;
 RA "A proline-rich protein, verpulin, involved in cytoskeletal
 RT organization and cellular growth in the yeast *Saccharomyces*
 RL cerevisiae".
 RL Mol. Microbiol. 10:585-596(1993).
 [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN-S288C / AB972;
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
 RA Favello A., Fulton L., Gattung S., Greco T., Kirsten J.,
 RA Kuwaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
 RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
 RA Mardis E., Meneses S., Miller N., Nhan M., Pauley A., Peluso D.,
 RA Rifken L., Riles L., Raich A., Trevaskis E., Vignati D.,
 RA Wilcox L., Wohlman P., Vaudin M., Wilson R., Waterston R.;
 RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases
 CC -1- FUNCTION: INVOLVED IN CYTOSKELETAL ORGANIZATION AND CELLULAR
 CC GROWTH. MAY EXERT ITS EFFECTS ON THE CYTOSKELETON DIRECTLY, OR
 CC INDIRECTLY VIA PROLINE-BINDING PROTEINS (E.G. PROFILIN) OR
 CC PROTEINS POSSESSING SH3 DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib.ch).
 CC -----
 CC
 DR EMBL; Z26645; CAA81388.1; -;
 DR EMBL; U19028; AAB67263.1; -;
 DR PIR; S39626; S39626.
 DR SGD; S0004329; VRPL.
 KW Cytoskeleton; Repeat.
 FT DOMAIN 5 14
 FT 239 245 POLY-PRO.
 FT DOMAIN 349 357 POLY-PRO.
 FT DOMAIN 396 406 POLY-PRO.
 FT DOMAIN 424 431 POLY-PRO.
 FT DOMAIN 462 468 POLY-SER.
 FT DOMAIN 704 708 POLY-PRO.
 FT CONFLICT 308 308 P -> R (IN REF. 1).
 FT CONFLICT 350 350 A -> R (IN REF. 1).
 FT CONFLICT 689 689 V -> E (IN REF. 1).
 FT CONFLICT 710 817 PSTMDGTGNSPSKSLKORLESTGSGTLQKHNTHTNOPDV
 FT DVGRTTGGNSITGAGSKRFRITDTSRFRKNTNVSGMPPL
 FT RPRKRTKLYPSGSSVPLDITLFT -> HIRITPLPLA
 FT FVKTLNNGYFLQVDRRCNTSITIRIQIMQLM (IN REF.
 FT 1).
 SQ SEQUENCE 817 AA; 82593 MW; 24C752D5B1CA1C8 CRC64;
 Query Match 3.28; Score 201; DB 1; Length 817;
 Best Local Similarity 20.28; Pred. No. 0.0099;
 Matches 154; Conservative 85; Mismatches 273; Indels 250; Gaps 34;
 OY 6 PGAPPLGAMPYDPDPSASEK---TSSLSPGLTNSGDSEFTTSAIIASVKEQLQ 61
 DB 150 PSSSAP---PLPDPSSAAPPIPIVPSPPAPPLPLSASAPKVPQNPMPVRAHRS 205
 OY 62 PERLTRELEAEROIIVASOLERCKLESETGSSMSASAEQFOWOSQDGKDIEDELTTGL 121
 DB 206 HQRKSSNLSLP-SVASAPPLPASLPTHSN----- 234
 OY 122 ELVDSCLNSLQSGILDPODYSTGERPSLSQSALQLNSKPEGSFOVPASHSQTLALG 181
 DB 235 -----PQAPPPPPPTPTIGLDSK---NIKPTDNAVSPSS----- 265

OY 182 ETTSPQLPARG-----TQARATGGSFSGTGTS-----RAGHLAGEP-----APP----- 221
 DB 266 ----SEVPAGGLEPLAEINARSRSEGAEGVSTKIQTENHSSPQPPPLPSAPPIPSH 321
 OY 222 -PPPPPEPFAASLGSASFILP-----DAP-----PAAAAALYYSSSTLPAPPRGGSPL 269
 DB 322 APPLPPTAPPPPSLPVTSAPKKAISAPAPPPPLPAAAMSSASTNSVATPPLPAPPL 381
 OY 270 AAPQGSPTKLRGSGAPGATYTAAPRGSSPKQSSRLAKSTGSPINIVYSSAGLSPI 329
 DB 382 PNTTSVPNPK---ASSMP-----APPPPPPPPGASTSSALS---ASSIPLAPL 425
 OY 330 RVTSPTVOSTSSPIHQLSSTIGTATLSPKRLVHNSQSHSOELVATATL----- 385
 DB 426 PPPPPSVATVPSAP-----PPPTLITNKS-----ASSQKSISSSSSSANTPPGP 475
 OY 386 -----QRPGLAAGRASYSQGH-----LGEPLALQSP 416
 DB 476 LPPLAEIQKKRDRFVVGDTGYTDQKQEDYIGSSKQDNVBPSPISPINPKQSSONG 535
 OY 417 EHHIDPIEDRYQK-----PPKRSLSQSGDPLPA-----HT 450
 DB 536 MSFLDEI-BSKLIKQTSNANFAPPH--TDMAPPLPSAPPPPTSLPPTASGDHT 592
 OY 451 GT-----YRTSAPSSPG-YDSVPLQRTGSOHGPNAAATFORASVAAAGPASN- 499
 DB 593 NDKSEVYLGKKAKKAPALGHPVPPVPLSDSKNNVPASL---LHDVLPSSNLEPK 649
 OY 500 -ADVPQLOVCPSESP-----YKSGPALPEEG-TLARSPLDSIQKDPREFGRDPE 551
 DB 650 PSPPVAAAPPLPFTSAPSLPOQSVSTSISSPPPVAPTLVSTRETESISKNPYK-SPPPP 708
 OY 552 LPEVIMLOHQFSPVQSNAAVYLQHLCEFGDNKIKAEIRROGGIQLVLDLDRM----- 605
 DB 709 SPSTMDTGTSNPS-----KNIKQRLSTGG-----STLQKHNTHTNQ 747
 OY 606 -TEVHRSAGALRLNVYGRAN-----DNNKIALKNCGIP 639
 DB 748 PDVDVGRYTTIGGSNSLIVGAKSGNERIVIDSRRKMTNVSGMP 789
 RESULT 9
 SMFL_HUMAN STANDARD; PRT; 1902 AA.
 ID 014497; O9UPZ1.
 AC 01-OCT-2000 (rel. 40. Created)
 DT 01-OCT-2000 (rel. 40, Last sequence update)
 DT 01-OCT-2000 (rel. 40, Last annotation update)
 DE SWI/SNF-RELATED, MATRIX-ASSOCIATED, ACTIN-DEPENDENT REGULATOR OF
 DE CHROMATIN SUBFAMILY F MEMBER 1 (SWI-SNF COMPLEX PROTEIN P270) (B120).
 GN SMARCE1 OR CIOBF4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND MUTAGENESIS.
 RX MEDLINE=20221560; PubMed=10757798;
 RA Dallas F.B., Pachione S., Wilsker D., Bowrin V., Kobayashi R.,
 RA Moran E.;
 RT "The human SWI-SNF complex protein p270 is an ARID family member with
 RT non-human specific DNA binding activity".
 RT Mol. Cell. Biol. 20:3137-3146(2000).
 RN [2]
 RP SEQUENCE OF 1-1175 FROM N.A.
 RX MEDLINE=98094256; PubMed=9434167;
 RA Takeuchi T., Chen B.-K., Qiu Y., Sonobe H., Ohtsuki Y.;
 RT "Molecular cloning and expression of a novel human cDNA containing CAG
 RT repeats".
 RL Gene 204:71-77(1997).
 RN [3]
 RP SEQUENCE OF 1-1132 FROM N.A.

ALTERNATIVE SPLICING.
 RX MEDLINE-94061824; PubMed-8242607;
 RA Oshima M., Sugiyama H., Kitagawa K., Taketo M.;
 RT "APC gene messenger RNA: novel isoforms that lack exon 7.";
 RL Cancer Res. 53;5589-5591(1993).
 CC -1- FUNCTION: TUMOR SUPPRESSOR. ALLOWS THE RAPID TURNOVER OF BETA-
 CC CATEININ. APC ACTIVITY CORRELATED WITH ITS PHOSPHORYLATED STATE
 CC ALLOWS THE DOWNREGULATION OF CYTOPLASMIC BETA-CATEININ (BY
 CC SIMILARITY).
 CC -1- SUBUNIT: FORMS HOMODIMERS AND ASSOCIATES WITH CATEININS (BY
 CC SIMILARITY).
 CC -1- ALTERNATIVE PRODUCTS: 4 ISOFORMS; 1 (SHOWN HERE), 2, 3 AND 4. ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN LIVER, SPLEEN, KIDNEY, HEART,
 CC LUNG, BRAIN, STOMACH, INTESTINE, TESTIS AND OVARY.
 CC -1- PFM: PHOSPHORYLATED BY GSK-3B (BY SIMILARITY).
 CC -1- SIMILARITY: CONTAINS 7 ARM REPEATS.
 CC -----
 CC THE SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M88127; AAB59632.1; -;
 CC EMBL: U02937; AAA03443.1; -;
 CC DR HSSP; Q02248; ZBCT.
 CC DR MGD; MG1:88039; APC.
 CC DR InterPro; IPR000225; -;
 CC DR Pfam; PF00514; Armadillo_seg. 4.
 CC DR PROSITE; PS50176; ARM_REPEAT. 1.
 CC DR Anti-oncogene; Phosphorylation; Alternative splicing; Repeat;
 CC Colled coil.
 KW DOMAIN 1 61 COILED COIL (POTENTIAL).
 FT 125 245 COILED COIL (POTENTIAL).
 FT DOMAIN 1 728 LEU-RICH.
 FT REPEAT 451 493 ARM 1.
 FT REPEAT 503 545 ARM 2.
 FT REPEAT 546 589 ARM 3.
 FT REPEAT 590 636 ARM 4.
 FT REPEAT 637 681 ARM 5.
 FT REPEAT 682 723 ARM 6.
 FT REPEAT 724 765 ARM 7.
 FT DOMAIN 739 2834 SER-RICH.
 FT DOMAIN 1130 1136 ASP/GLU-RICH (ACIDIC).
 FT DOMAIN 1556 1575 ASP/GLU-RICH (ACIDIC).
 FT DOMAIN 1864 1891 HIGHLY CHARGED.
 FT VARSPLIC 243 276 MISSING (IN ISOFORM 2 AND ISOFORM 4).
 FT VARSPLIC 310 410 MISSING (IN ISOFORM 3 AND ISOFORM 4).
 FT VARIANT 120 120 T -> A (IN STRAIN CAST/EI).
 FT VARIANT 493 493 V -> I (IN STRAIN CAST/EI).
 FT VARIANT 797 797 Y -> F (IN STRAIN CAST/EI).
 FT VARIANT 1330 1330 A -> T (IN STRAIN CAST/EI).
 FT VARIANT 1618 1618 A -> S (IN STRAIN CAST/EI).
 FT VARIANT 2294 2294 G -> A (IN STRAIN CAST/EI).
 FT VARIANT 2496 2496 H -> Q (IN STRAIN CAST/EI).
 FT VARIANT 2523 2523 T -> A (IN STRAIN CAST/EI).
 FT VARIANT 2813 2813 T -> S (IN STRAIN CAST/EI).
 SO SEQUENCE 2845 AA; 311086 MW; 145CAV3CF570A499 CRC64;

Query Match 3.18; Score 200; DB 1; Length 2845;
 Best Local Similarity 17.98; Pred. No. 0.053;
 Matches 225; Conservative 170; Mismatches 428; Indels 434; Gaps 48;

OY 23 SASERTSLSP-----GLNTSNGDSETE-TTSAILASYKEOF-----LQ 61
 DB 100 SVSSSGSCSPVPMKSPFRRTFVNSRGSTGLLEKERSLLDLDKEXEKDWYAAQ 159
 OY 62 FERLIRELEA-----EROIVASOLE-----RCKLIGSETGSMSSMSAEQFOQMS 106
 DB 100 SVSSSGSCSPVPMKSPFRRTFVNSRGSTGLLEKERSLLDLDKEXEKDWYAAQ 159
 OY 62 FERLIRELEA-----EROIVASOLE-----RCKLIGSETGSMSSMSAEQFOQMS 106

DB 160 LQNTFRKIDSLPLETENFSIQDTMTROLEVEARQIRAMEEQLGTCODEMERRAQRIRARI 219
 OY 107 QDGCQDIEDDELTTGGLVDSCIRLSQESGIIDPDYISGEBPRLSSALQNKPGSF 166
 DB 220 QQIEEDI-----LRRLQDS-----QAAERSS-----QSRHDAS 252
 OY 167 QPVASHNSQTLALGETTSPOLPARGTARATGGSFSGSTSRAGHLGAPPPAPPPPP 226
 DB 253 HEAGRQHECHGVAESENTAAS-----SSGSPATRVDH----- 284
 OY 227 REPFAPIGSAFHPDAPPAALALYSSTLPAPPPGSPAPAGSGPTKIQRGSA 286
 DB 285 -----ETASVLSGSHSAPRLTSHLG-----TKVEMYS- 315
 OY 287 PEGATYAAPRGSSPSPSPRLAKSYSTSPITIVVSGSLPIRYTSPPTVOSTISSPI 346
 DB 316 -----LLSMGLTHDKDMSKRTLLAMSSSDSCISMRGCGPILI----- 355
 OY 347 HOLSTTGTATLSPTKRLVNASQYSKHSELVATATLQRPGLAAGSASYSQHG- 404
 DB 356 -----QLHGNDKDS-----VLGNSRSGKEARARAS-AALHNII 389
 OY 405 HLGPE-----LRALQSPHHIDPIEDRYVQKPPMSLSQSGDDPLPPAHGTYYR 454
 DB 390 HSQPPDKRGRREIRYLHLEQIRAYCETCWE--WQEAHEQGMDDK-NPMP----- 437
 OY 455 TSTAPSPGVDSVPLQRTGSGHGPNAAAATPQRASVAAAGPASVADPYROLQCPSVES 514
 DB 438 -----APYDH-----QICPNV- 448
 OY 515 PYKSGPALPREGTLARSPSISIQKP-REKGRWDPPELPEYIOMQ-----HQ 562
 DB 449 -----CYLMKLSFDEEHRHANNELG-----GLQALAEHLQYDCMYGLTNDY 491
 OY 563 PFSVQSNAAAYLOHLQCFGNKTKAEI-RRQGIOLLYOLLHRMFEVRSAGLARNLY 621
 DB 492 SVTLRAGAMALTNLTFFGVANKATLCGMRALVQALQSESDLOQVATVARNLSW 551
 OY 622 GRANDNKTALKNCGIPALVRLKRTDLEIRELVTVGLNMLS-CDALKNPI--IODA 678
 DB 552 -RADVASKTLEKGVSKALMECALEVKESTLKVSLAMLSHCHENKADICAVGDA 610
 OY 679 LAVLTNAVITPHSGWENSPLODKRIQLHSQVLRNATGCLRNVSS--AGEEARRRRE 735
 DB 611 LAFVLGTLY-----RSQTNLTALIESGGIILNVSLSLATVEDHRIOLRE 656
 OY 736 CGGLTALLIYQSLAGSSEIDSKTYENCVCILNLSYRLAETSGQGMGDELDELGLC 795
 DB 657 NNCLQTLQHLKSHSL-----TIVSNAGTLMNLISAR----- 688
 OY 796 GEANGKDAESSGCWKKKKKKKSDQMDGVGPLPDCAPPKGOMLWPSIVKPYLTLLS 855
 DB 689 ---NPKDQA-----LMDGAVS-MLKNLI 709
 OY 856 ECSNPDTLGAAGALONLAA-----GSMKSVYIRAAVKEKGLPLVELL 901
 DB 710 HSKHKMIAMGSAALALNLMANRPARKYKANINSPSSLPSTH---YKQKRLAEIDAQ 765
 OY 902 -----RIDNRRVCAVATLARNMALDVNKKELIGYAMRDLVHRLPGNNSNTASKA 954
 DB 766 HLESTEDNDN-----ISPKASHSKQHKHONLIGDYAF-----DANHDSRDN 811
 OY 955 MSDDTAVACCTLHEVITKNMENAKALRDAGGIEKLVGISKSG----- 998
 DB 812 FNTGNMTVLSPLYNTVPLSSSSSSRGLSDSRSLERERIGLSAYHPTTENAGTS 871
 OY 999 DKHSPPVKAASOVLSNMQY-----RDLRSLYKKDGSOYHFAVSSSTTRDRQRYX 1052
 DB 872 SKRGLOITTTAAQIAKVMEEVSAIHTSDDRS---SASTPEHCVAADRSARRSASHT 928
 OY 1053 SSRTPSISVPRVSPNNRSASAPASPREMISLKEKRTDYCTGNSNTYAGAKGEHTSRDA 1112
 DB 929 HSNVTYNT--KSENSNKTCSMPA--KVEYKRSSNDLSNVTSSDGIKKRQ--MPS 980

OY 1113 MTQONTISILYRNSYCAPADIKHNOVSAQPYOEPSRKDYETOPFONSTRNDE 1169
 DB 981 VESYSEDESKF-CSYGOYPADLAHKIHSANHMD-NDGLDLPINSLKYSDE 1032

RESULT 11
 IE18_PRIVIF STANDARD; PRT: 1461 AA.
 ID IE18_PRIVIF
 AC P11675;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 01-FEB-1994 (Rel. 28, Last annotation update)
 DE IMMEDIATE-EARLY PROTEIN IE180.
 GN IE.
 OS Pseudorabies virus (strain Indiana-Funkhauser / Becker) (PRV).
 OC Viruses; dsDNA Viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Varicelloviruses.
 RX NCBI_TaxID=31523;
 RN [1]
 RA SEQUENCE FROM N.A.
 RX MEDLINE=89315207; PubMed=2546124;
 RA Cheung A.K.;

RT "DNA nucleotide sequence analysis of the immediate-early gene of
 pseudorabies virus."
 RL Nucleic Acids Res. 17:4637-4646(1989).
 RN [2]
 RA Cheung A.K.;

RP REVISIONS.
 RL Submitted (NOV-1989) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: THIS IE PROTEIN IS A MUTIFUNCTIONAL PROTEIN CAPABLE
 OF MIGRATING TO THE NUCLEUS, BINDING TO DNA, TRANS-ACTIVATING

OTHER VIRAL GENES, AND AUTOREGULATING ITS OWN SYNTHESIS.
 CC -1- SUBCELLULAR LOCATION: NUCLEUS OF INFECTED CELLS.
 CC -1- PTM: A LONG STRETCH OF SERINE RESIDUES MAY BE A MAJOR SITE OF
 PHOSPHORYLATION.

CC -1- SIMILARITY: BELONGS TO THE ICP4/IE140/IE180 FAMILY.
 CC CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 or send an email to license@sdb.ch).

CC EMBL; X15120; CAA33214.1; .
 DR PIR; S04713; EDBE1F.
 KW Early protein; Transcription regulation; Trans-acting factor;
 KW DNA-binding; Phosphorylation; Nuclear protein.

FT DOMAIN 390 405 POLY-SER.
 FT SEQUENCE 1461 AA; 149833 MW; 7F31E7ABE403B208 CRC64;

Query Match 3.1%; Score 197; DB 1; Length 1461;
 Best Local Similarity 20.9%; Pred. NO. 0.03;
 Matches 177; Conservative 99; Mismatches 347; Indels 224; Gaps 37;

OY 1 MFARKPPGAPLAMPVPDPSPASSEKTS-----LSPGLTNSGDSSETTSAIILAS 54
 DB 18 LIAAAAAAEEGIGASPGGSGSRRSSGDELLFGRG-GLFSDDAAEAE-AAVILAA 74
 OY 55 VAEQELQFRLRLREAEHQIYASOLERCKLGESESSSSSSAAEPOFOWOSODGOKDIE 114
 DB 75 A-----AGATRRPPPPSAOORHARRGS--GEIVLDDDEDEE-----E 110
 OY 115 DELTGLLELVDCISLQSSG-----ILDPDYS--TGERPSLLSQAOLNSKPEG----- 164
 DB 111 DE--FSPAPAGSPGRALHSGHGLVIGFPRASGSPPPPALAAAEAGAPGPGRS 168
 OY 165 --SFOYPASYHSNQTALGETTPSQLPARGTQARATGOSFSOGTTSRAGHLAEPAPPP 222

DB 169 SPASAPASSSSSGSSGSSGSPGSAAPRRMSPAR-----GDPVGERP 211

OY 223 PPPPREPAPSLGSAFHPDPAPAAAAALYSSSTLAPRGSGPLAPGGSPTKLQR 282

DB 212 AARPRP-----APPAOAAVAAPARRGAPASPAAG-PVSAF-----G 251

OY 283 GGAPEGATYAAPRGSSPKQSPRLAKSYSTSPINIVYVSGSLPIVTSPTVQSTIS 342

DB 252 GGGARSG---GGDRGRHHQHREPL-----LDEPAAARRLDP---RPLARSEVS 295

OY 343 SSPHQLSSTIGTATVTLSPYKRLVHASQYKSHSQELVATLQR-----DGLAAG 394

DB 296 SNPNSSSSST--TVAAVEPVARGPEKDEDGLAGD--GGAFLQRPARRRAGGALRG 351

OY 395 SRASYSQHGLHPELRRLQSPHHIIDIYEDRYQKPMKSLQSGODPLPAPHTGYR 454

DB 352 -RGFSSSSSGGSDSLSPARSPAPRAAAAAAARRSASSSSSSSS----- 398

OY 455 TSTAPSPGVN-----SVPLQRTGSGQHP-----ONAAATFORASYAAGPASN 498

DB 399 SSSSSSSGGEDEGVPRGAPLAPAGPPSPAPAPAPASASATSSSAAPAPAE 458

OY 499 YADPTROLQYCPD-----VESPTKSGPALPPEGT--LARSPTDSTQK---DPREF 545

DB 459 PARPRRRRRRTNNHLSLMADGPPTDGPPLTLPGEPPGSDPADGRVRYGAGDSREG 518

OY 546 GWRDELEPEVIONLQHOPPSVQSNAAATLOHLCFGDNKIKAEIRGGIQLVLDLHRM 605

DB 519 LMDED-----VRQAAARKRAAAGPVPTIPRM--GDSKQHE----- 554

OY 606 TEVRHSACALRNIVYGRANDNKAINKCGIPALVRLKTTDLIRE----- 655

DB 555 -----ALVRLIYSGAGEAMSWLQN---PRM-----QANDQAFNQCQRVHAPH 596

OY 656 ----LYTGVLNLSSCDALKPIIQDALAVLTNAVITPHSGWENSPLODDRKIQLHSSQV 711

DB 597 GHGSFTTSV-----TPPLPHIGDMAADPLMAAPHA--VSAAVMSRRYRTQKTFI 647

OY 712 LRNATGCLRNVSSAGEARRRRCDDGLTALVYIOSALGSSP---IDSKVENCICIL 768

DB 648 LOSLRRAADAAAYGRRADPPAGE--ATVEALCARVRAAPAAAPGVRRELADACVLAC 705

OY 769 RNLSTRL 775

DB 706 RGVLERL 712

RESULT 12
 ID YM96_YEAST STANDARD; PRT: 1140 AA.
 AC 004893;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HYPOTHEMETICAL 113.1 KDA PROTEIN IN PRES-FET4 INTERGENIC REGION.
 GN YMR317W OR YM9924.09.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 RX NCBI_TaxID=4932;
 RN [1]
 RA SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Churcher C.M., Louis B.J., Barrell B.G., Raeburn M.A., Walsh S.V.;
 RL Submitted (NOV-1995) to the EMBL/Genbank/DBJ databases.
 CC -1- DOMAIN: CONTAINS MANY SER/THR-RICH DOMAIN AND REPEATS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>

or send an email to license@isb-sib.ch.

CC or send an email to license@isb-sib.ch.
 DR EMBL; Z54141; CA930835.1;
 DR SGD; S0004936; YMR317W.
 KW Hypothetical protein; Repeat.
 SQ SEQUENCE 1140 AA; 113070 MW; 0153EBCA24FE5427 CRC64;

Query Match 3.1%; Score 195; DB 1; Length 1140;
 Best Local Similarity 19.4%; Pred. No. 0.029;
 Matches 242; Conservative 177; Mismatches 505; Indels 322; Gaps 52;

20 QPSASAKTSLSPGLNTSNGDSEETETTSAILASYKEDQLQ 62
 128 QLSSTETSTISSAQTSSPOTSSNGGSSSEPLGKSSVLETTAKSSDTTAVTSST 187
 63 ERLRELEAEQIYASOLEKCLGSEFSMSMSMAEQQOMOSODQKIDELTGTGLE 122
 188 TLTTL-DYSSSKISSGSANTSVGTISDAK-----E 218
 123 LVDSCLRSLOESGILDPODYSTGERPRLSQALQNSKEGSEFOYPASHNOTALGE 182
 219 VFSS-----STDVSSLLS-----STSSPASSTISETLPFSS 250
 183 TTFS-QLPARGTQARATGGSFGCTTSRAGHLAGEPAPPPPPPPPPPPPPPPPPPP 240
 251 TLTITSSVSSSEAPSAITSSVSSSEASSSTSSVSE-APLALTSSVSSSEAPSTSSVS 309
 241 PDAPPAALALYSSSTLPAPPRGSPPLAAPQGSPTKLQSGSAPEGATYAAPRGSSP 300
 310 SEASSSTSSVSSSEISSTSSVSSSEAPLA-----TSSVSEAPSSSTSSVSEISS 362
 301 KQSPRLAKS-YSTSSPLNIYSSAGLSPIRTVSPPTVOSTISSPFIHOSSRTIGTATL 359
 363 TTSSSVSEAPLATSSVSSSEAPSTSSVSEAPSTSSVSEAPSTSSVSEAPSTSSVS-SET 420
 360 SPTRLVASEQYKSHSOELATATL--GRPGSLAA-GSRASYSOHLGPELRALQS 415
 421 SSTSSVSSSEVSS-----ATSSIVSEAPSAISSLASRLFSKNTVSTSLVATEA 473
 416 PEHHIDPIEDRYOKPRLMSLSQSGDPLPRAHTGYRTSTAPSSGCVDSVPLQRTGSO 475
 474 -----SSVTSLSRPSSETLANSITTESLS 498
 476 HGPNMAAATFORASYAGPASNATADPYRQOLCPSEVSPKSGPALPREGTLARPSI 535
 499 TGVNVTSTTSSAASSTIGSVSSSSNMATSKTSTSSLSKSVLPFGNSVTYTTSPSA 558
 536 D-SIOKDPREFGWRDPELVYIQLHOFPSVQSNAAAYIQLHCFGDNKIKAEIRROGI 594
 559 SISLASPLPSVSD-----ITSSSEASSISNLAS--SAPSDNNSTI-----ASA 602
 555 QLLVDLDHRTTEVHRKACGL-RNLVYGRANDDNKIALKNCGI--PALVRLLRKTTDL 651
 603 SLIV-----TKTKNSVSVSSIVSITSEETNESNLATSTSLSKMKAFAISLSTNAT 655
 652 EIRELVTGVNLNLSGCDLKPITODALAVLT-NAVITPHS--GM-----ENSLQDRK 703
 656 SASNPVTGFSMSHSTVITPGSTSSASLAISIVYSSSLAISTPSPSSPTT----- 711
 704 TQLHSSQVLRNATGCLRNVSSA-----GEEARRRMRCDGLTALLVYIQSALGSSEID 757
 712 -----STLVTSAPSVSSMTSAPFINNSTARSPPASTSTESTSSISSVPLASGDT 767
 758 SKTYVNCICILRNLSYRLAETSGQOHMGTELDGLGCEANGDAESGCGWKKKKKK 817
 768 SSLAAH-----NLITFSAPTSSAQIV-----SKSTSSSTI----- 798
 818 SODQNDGVGPLDCAEPKGIQMLHPSIVKPYLLTSECNPTLLEGAAGALON----- 872
 799 -----LVMP-----RIDRSGNNSPASHIAISLNNKTTFV 827
 873 --LAAGSKMSVYIRAAVRKEKGLPIVELLRID--NDRVVCAYATALRNALDV---RN 925

DB 828 SLSSTS-----AHRNIFNSVTLATAKQIEFLNIVWCNSPFPNNYNTKTVIVSRE 879
 OY 926 KELICKYMRDLVHRLPGC--NNSNNPASKMSP-DVYTAVCCLLHEVITKNNENALR 982
 DB 880 TTAID-----TVNISCSCGCTKRNKSTLTITIDASTVTTCEKEV----- 921
 OY 983 DAGTEKLVGISKSGD--KHSPKVKASQVLSNMQYRDLRSILY--KKDGSQYHFA 1038
 DB 922 -----TSTISGDAEHTTSTKISNFETSTPSEFKMKISQETKRAKPSSEIVRS 971
 OY 1039 SSSITDRORPYSRRTPSPVAVSPNNRSASAPSPREMISLKERKTDYECTGSNAT 1098
 DB 972 SSSFEK-----TSPITKASP-STSPSEKKAAGNVSATNAPSPSTSPSESGTGSTSV 1023
 OY 1099 YHAKGEHTSRKDMAT---AQTGISTITRNSYCAPADIDHNOVASQAPVQEPSRKDYE 1155
 DB 1024 -EGAKSKSTKSEGVSTTKARKNT-STVAKSSTESP--IGRGTLETITIVS-SQKSL 1076
 OY 1156 TYQPFQNSTRNDESEFED--QVHRRPASE---YTMHLGKSTG 1195
 DB 1077 TSO-ISSSTEKYNKSTTTPTAIHGTSSAKOSTIYVSTAKENTG 1121

RESULT 13

APC_RAT ID APC_RAT STANDARD: PRT: 2842 AA.
 AC P70478;
 DT 01-OCT-2000 (Rel. 40, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE ADENOMATOUS POLYPOSIS COLI PROTEIN (APC PROTEIN).
 GN APC.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FISCHER 344/N; TISSUE=Brain;
 RX MEDLINE=96116966; PubMed=8563176;
 RA Toyota M., Ushijima T., Kakiuchi H., Watanabe M., Imai K., Yachi A., Sugimura T., Nagao M.;
 RT cDNA Cloning of the rat APC gene and assignment to chromosome 18.;
 RL Mamm. Genome 6:746-748(1995).
 RP [2]
 RP MUTAGENESIS.
 RC STRAIN=SPRAGUE-DAWLEY, AND FISCHER 344/N;
 RX MEDLINE=95148647; PubMed=7846077;
 RA Kakiuchi H., Watanabe M., Ushijima T., Toyota M., Imai K., Welsburger J.H., Sugimura T., Nagao M.;
 RT Specific 5'-GGGA-3'->5'-GGA-3' mutation of the APC gene in rat colon tumors induced by 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine.;
 RL Proc. Natl. Acad. Sci. U.S.A. 92:910-914(1995).
 CC -1- FUNCTION: TUMOR SUPPRESSOR. ALLOWS THE RAPID TURNOVER OF BETA-CATENIN. APC ACTIVITY CORRELATED WITH ITS PHOSPHORYLATION STATE ALLOWS THE DOWNREGULATION OF CYTOPLASMIC BETA-CATENIN (BY SIMILARITY).
 CC -1- SUBUNIT: FORMS HOMODIMERS AND ASSOCIATES WITH CATENINS (BY SIMILARITY).
 CC -1- PTM: PHOSPHORYLATED BY GSK-3B (BY SIMILARITY).
 CC -1- SIMILARITY: CONTAINS 7 ARM REPEATS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL; D38629; BAA07609.1;
 CC DR HSSP; Q02248; 2BCI.

InterPro: IP000225; -
 DR Pfam: PF00514; Armadillo_seg; 4.
 DR PROSITE: PS0176; ARM_REPEAT; 1.
 KW Anti-oncogene; Phosphorylation; Coiled coil; Repeat.
 FT DOMAIN 1 728
 FT DOMAIN 1 62
 FT DOMAIN 125 260
 FT REPEAT 451 493
 FT REPEAT 503 545
 FT REPEAT 546 589
 FT REPEAT 590 636
 FT REPEAT 637 681
 FT REPEAT 682 723
 FT REPEAT 724 765
 FT DOMAIN 739 2831
 FT DOMAIN 1130 1155
 FT DOMAIN 1556 1575
 FT DOMAIN 1864 1891
 FT MUTAGEN 523 523
 SQ SEQUENCE 2842 AA: 310530 MW: 308286834E8F47 CRC64;
 C->R: IN AN IO-INDUCED COLON TUMOR.

Query Match 3.1%; Score 194.5; DB 1; Length 2842;
 Best Local Similarity 18.1%; Pred. No. 0.096;
 Matches 230; Conservative 166; Mismatches 424; Indels 451; Gaps 50;

23 SASEKTSLSLP-----GLTNSGDSSETE-TTSALISVKEOE-----IQ 61
 100 SVSRSGECSPPVPMGSPRRAPVNGSRETYLELEKERSLLADLDEEEKDKWYQAQ 159
 62 FERLTFRELEA-----EROTIVASOLE-----RCKIGSETGSSMSKSAEQQWQS 106
 160 LQULTRIDSLPTENFSLOTDMTRQLEYARQITAAEBQIGTQDEKRAQRIARI 219
 107 QDGKQIDELTGLVDSCLRSLOESGILDPQDYSTERPSLSQSALQINSKEGSP 166
 220 QQTEKDI-----LRVROLLOS-----QAABEABRS-----QSKHE----- 249
 167 QVPASVHNSQTALGETTSQLPARTQARATQSGTSQGTSGAGHLAPPEAPPPPPP 226
 250 --TASHEAEKQLEGGVASNSLATSGS-----GQSSAAKRVH----- 284
 227 REPFASLSAFLPAPPAALAAALYSSSTLPAPPPGSPPLAQQGSPTRKLRGSA 286
 285 -----ETAGVLSSSGHSHAPRLTSHLG-----TKVENYVS- 315
 287 PEGATTAAPRGSSPKQSPRLAKSVSTSSPINIVVSSAGLSPIRTVSTPTVSTSSPT 346
 316 -----LLSMIGTHDKDMRSTLLAMSSQDSCISMROGCLPLLI----- 355
 347 HOLSTIGVYATISPTKRLVHASEQYSKSHSOELVATATIQRFCSLAAGRASYSOHC-- 404
 356 -----QLHNSDKDS-----VLGNSRSGSEKARARAS--AALHNII 389
 405 HLGPE-----LRAQSPENHIDPIEDRYVOKPRMSLSOSOGPPLPAHTGYR 454
 390 HSGPDDKRRRLRLVHLLEQIRACETCWE---WQEAHQGDOK-NPMP----- 437
 455 TSTAASSPEVDVPLQRTGSSQHGPNAAAATFORASVYAGPANSYADPYROLQYCPVES 514
 438 -----APVEH-----QICPAV-- 448
 515 PYKSGPALPREGTLARSPSIDIQKP-RFEGWMDPELPEYIOMQ-----HQ 562
 449 -----CYLMKLSFDEEHRHANNELG---GLQALAEILLQVDCENHGLTDHY 491
 563 PPSVOSNAAYLQHLCEFGDNKIKAEI-RRQGIQLLVLDLDRHMEVRSAGCALRNLVY 621
 492 SYTLRRYAGMALTNLTFQGVANKATLCSMKCGMRALVADQKSESDIQVIAVLRNLVSW 551
 622 GRANDNKAALNCGGIPALVRLKTTDLTRELVTGVNLVSS-CDALMPT--TODA 678
 552 -RADVNSKKTTLREGVSVKALMECALELVKKESTLKSALMNLMSAHTENKADICAVGDA 610

679 LAILTNAVLIIPHSWENSPLODDRKIOLHSSQVLRNATGCLRNVSS---AGEFARRBRE 735
 611 LAFIVGTILY-----RSQNTLAIISGGGILRNSSSLIATEDHRQILRE 656
 736 CDGLTDLALIVYQASLGSSSEIDSKTVENCYCLRNLSYRLAETSGQHMGTDELGLLC 795
 657 NNCLQTLLOHLKHSLSL-----TIVSNACGTLNLSAR----- 688
 796 GEANKKDAESSGCKMKKKKKKSDQMDGVPLPDCAEPPKGIOMLHPSTVKKYLTLLS 855
 689 ---NPKDEA-----LWDMGAVS--MLKNLI 709
 856 ECSNPDTLEGAALONLAA-----GSMKSVYIRAVKREKGLPIVELL 901
 710 HSKHMMIMGSAALRNLMANRPAYKANIMSPSSLSPLSH---VRKQALAEIDAO 765
 902 -----RIDNDRVCAVATLRNMLDVNRKELIGKYAMBDLVHRLPGNNSNMTASKA 954
 766 HLSETFEDNIDN-----LSPKASHRSKORHKQNLGYGVYFPASRH---DDNRSN- 811
 955 MSDDVTAVVCCTLHEVITRKNNENKALRDAGGIEKLVISKSG----- 998
 812 FNTGMMTVLSPLNTTVLPSSSSRGLSDSRSEKDRSLEREIGLSTYHSATENPGTS 871
 999 DKHSPPVKAASQVNLNSMAYRDLRLSLYKKDQWS-----QYHFAVASSSTTERDQRPYSS 1053
 872 SKRGQLSATGAQIAKWEVYSALHT--SQDDRSPASMAELHCAVE---ERTARRSSA 925
 1054 SRTPTISPVYRSPNNRKSAPAREMISLKEKRTIDYECTSNATY-----HGANGEH 1106
 926 SHT-----HPNTHNPAKSESSNRCTSMRYAKVEYKRSSNDLSNVTSDDYGKKGOM 977
 1107 TSKDAMTAONTGISTILYRNSYGAPAEIDIKHNOVSAQVPEPSPKRYETVQPPQNSTRN 1166
 978 KPVSYSYSEDEGKFC-----SYGQYPADLAKHHSANHMDNGELD-----TPIN 1024
 1167 YDESFEEDQVH 1177
 1025 YSLKYSDEQLN 1035

RESULT 14
 YHL1_EBV
 ID YHL1_EBV STANDARD; PRT; 660 AA.
 AC P03181;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE HYPOTHETICAL BHLF1 PROTEIN.
 OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Lymphocryptovirus.
 OX NCBI_TaxID=10377;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=84270667; Pubmed=6087149;
 RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
 RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seglin C.,
 RA Tuffnell P.S., Barrell B.G.;
 RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
 RL Nature 310:207-211(1984).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC EMBL: V01555; -; NOT_ANNOTATED_CDS.
 DR PIR: A03742; Q0BE3.

KW Hypothetical protein; Early protein; Repeat.
 FT DOMAIN 149 648 4 x 125 AA TANDem REPEATS.
 FT REPEAT 149 273 1.
 FT REPEAT 274 398 2.
 FT REPEAT 399 523 3.
 FT REPEAT 524 648 4.
 SQ SEQUENCE 660 AA; 66244 MW; 86DA1D67A37152A2 CRC64;

Query Match 3.0%; Score 194; DB 1; Length 660;
 Best Local Similarity 22.4%; Pred. No. 0.016;
 Matches 161; Conservative 48; Mismatches 271; Indels 240; Gaps 33;

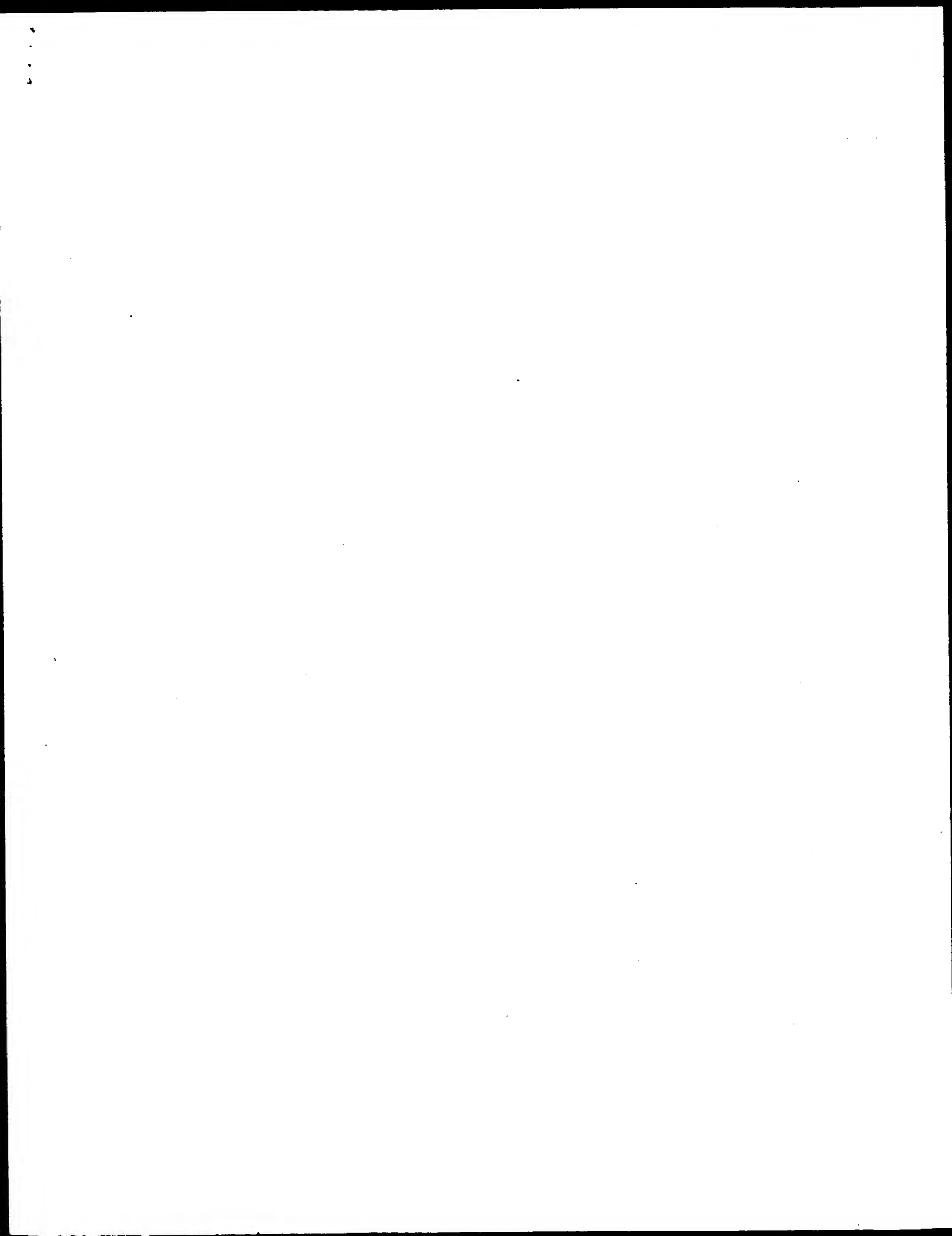
7 PGAAPLCAMPYPPDPSASEKTSLSPLGLTNSGDSSETTSAIIA-SYKEDLEQERL 65
 11 PRTPLPHCPPLPGAPDOOTRLPPGW-----GORTAPTOVGLADAASPELDQAS 64
 66 TRELEAEROIYASOLECKIGSETGSMSSEECFOWOSODGQKIDELTTLGLIVD 125
 65 GARGGGNRVAG--KGRCTAPRSQRTGPAE-----QADHANSPTG----- 108
 126 SCIRSLQESGILDPO-----DYSTGE-----RPSLSQ----- 153
 109 -----GCSDPQRSPRTQAGALGEGSAGISGRPPHAFQVOMGARNPGCPTW 159
 154 ---SALQNSKRPESFYQYPASYSNQTALGETTSPQLPARGTOARTGSGFSGTTSRA 210
 160 RRRGAGRGHPPGAGRGPTGGRAPAGCTPAAPG-----GGAIVSGATPHR 214
 211 GHLAGPEPAPP-----PPPREPAP-SLGSAPHLDPAPP-----AAAAALY----- 254
 215 ER--GSGADPAPPAARLPENQERLQDLAAAGRCAPGPPRTSGAAAGRTHRPPGCP 272
 255 -----SSSTLPAPRGSPPLAAPGGGSPTLQGRGSA--P 267
 273 RSARNPGCPRTWRRRRSGAQRCHPPGAGQRPSPGTPGAPGAPGAPGAGGAAVP 332
 288 EGATYAAPRGSPKPSPEKSLAKSYSTSSPITIVSSAGLSPIRTYTSPTVOSTISSPIH 347
 333 SGATPHPRGSGAPDP-----AAARLPPER-QERLQ----- 365
 348 QLSSTIGTYATLSPTRLVHASEQYSKHSELYATLQRPGLAAGSRASYSQSHGLG 407
 366 DLAAAGRCAPGPPRTSGAAAGR--THRRPGCPRSRANPGCPTWRRRS-GAQRCHPP 421
 408 P-----ELRALQSPHHIDPIYEDRVYQKP 433
 422 PGAGRPSPGTPGAPGAPGAPGAPGAGGAAVPSGATPHPRGSGAPDPAPPAARLP 481
 434 MR-----SLSOSQGDPL--PRAHTG--TYRTSTAP-----SSPGVDSVPLQRTGSQ 475
 482 EROERPLRQDLAAAGRCAPGPPRTSGAAAGRTHRPPGCRSARNPGCPTWRRRSQAQ 541
 476 --HGPONA-----AAATFORASVYAGPASNADYRLOLYOCPSVESYSK--S 519
 542 RGHPPPGAGRPSPGTPGAPGAPGAPGAPGAGGAA-----VPSGATPHPRGSG 593
 520 GPALPPECTLARSPIDSIGDPRFEGWDELPEVLOMLQH--QPPSYSNAAVLIQH 576
 594 GPADPP--AAARLP-----PER--QERLQDLAAAGRCAPGPPRTSGAAAGRTH 640
 RESULT 15
 SYJL_HUMAN STANDARD: PRT: 1575 AA.
 AC 043426; 043425;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE 01-OCT-2000 (Rel. 40, Last annotation update)
 DE SYNAPTOJANIN 1 (EC 3.1.3.56) (SYNAPTIC INOSITOL-1,4,5-TRISPHOSPHATE 5-
 DE PHOSPHATASE 1).
 GN SYMOL.
 OS Homo sapiens (Human).

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 CC NCBI_TaxID=9606;
 CC [1]
 CC SEQUENCE FROM N.A.
 CC TISSUE=Cerebellum;
 CC MEDLINE=96088905; PubMed=9428629;
 CC Haffner C., Takel K., Chen H., Ringstad N., Hudson A., Butler M.H.,
 CC Salcini A.E., Di Fiore P.P., De Camilli P.;
 CC "Synaptojanin 1: localization on coated endocytic intermediates in
 CC nerve terminals and interaction of its 170 kDa isoform with Eps15";
 CC FEBS Lett. 419:175-180(1997).
 CC - FUNCTION: INOSITOL 5-PHOSPHATASE WHICH HAS A ROLE IN CLATHRIN-
 CC MEDIATED ENDOCYTOSIS.
 CC - CATALYTIC ACTIVITY: D-MYO-INOSITOL 1,4,5-TRISPHOSPHATE + H(2)O =
 CC D-MYO-INOSITOL 1,4-BISPHOSPHATE + PHOSPHATE.
 CC - ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG ISOFORM/SYNAPTOJANIN-
 CC 170 (SHOWN HERE) AND A SHORT ISOFORM/SYNAPTOJANIN-145; ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC - TISSUE SPECIFICITY: CONCENTRATED AT CLATHRIN-COATED ENDOCYTIC
 CC INTERMEDIATES IN NERVE TERMINALS. THE LONG ISOFORM IS MORE
 CC ENRICHED THAN THE SHORT ISOFORM IN DEVELOPING BRAIN AS WELL AS
 CC NON-NEURONAL CELLS. THE SHORT ISOFORM IS VERY ABUNDANT IN NERVE
 CC TERMINALS.
 CC - DOMAIN: BINDS TO EPS15 (A CLATHRIN COAT-ASSOCIATED PROTEIN) VIA A
 CC C-TERMINAL DOMAIN CONTAINING THREE ASN-PRO-PHE (NPF) REPEATS.
 CC - DOMAIN: THE C-TERMINAL PROLINE-RICH REGION MEDIATES BINDING TO A
 CC VARIETY OF SH3 DOMAIN-CONTAINING PROTEINS INCLUDING AMPHIPHYSIN,
 CC SHP4 AND GRB2.
 CC - SIMILARITY: IN THE CENTRAL SECTION; BELONGS TO THE INOSITOL-1,4,5-
 CC TRISPHOSPHATE 5-PHOSPHATASE FAMILY.
 CC - SIMILARITY: CONTAINS 1 SAC1 DOMAIN.
 CC - SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AF009040; AAC51922.1; -
 CC EMBL: AF009039; AAC51921.1; -
 CC MIM: 604297; -
 CC DR InterPro: IPR000300; -
 CC DR Pfam: PF00783; IPIC: 1.
 CC DR PROSITE: PSS0102; RRM: 1.
 CC KW Hydrolyase; Alternative splicing; Repeat; Endocytosis; RNA-binding;
 CC Multigene family.
 CC FT DOMAIN 1 499
 CC FT DOMAIN 500 899
 CC FT DOMAIN 902 971
 CC FT DOMAIN 900 1575
 CC FT DOMAIN 1033 1036
 CC FT DOMAIN 1108 1113
 CC FT DOMAIN 1126 1129
 CC FT DOMAIN 1487 1490
 CC FT DOMAIN 1540 1546
 CC FT DOMAIN 1396 1419
 CC FT REPEAT 1396 1398
 CC FT REPEAT 1406 1408
 CC FT REPEAT 1417 1419
 CC FT VARSPLIC 1306 1311
 CC FT VARSPLIC 1312 1575
 CC SQ SEQUENCE 1575 AA; 173345 MW; 50646F6CC043B9E7 CRC64;
 VTNIGI -> QOEPG (IN SHORT ISOFORM).
 MISSING (IN SHORT ISOFORM).
 Query Match 3.0%; Score 194; DB 1; Length 1575;
 Best Local Similarity 24.1%; Pred. No. 0.049;
 Matches 137; Conservative 56; Mismatches 221; Indels 154; Gaps 29;

QY 105 QSDGQKIDELTGLGLVDSIRSLQESGIL-----DPQDSTGERPSL 150

```
Db 969 KSPDWIKNLEEMMS--LEKISIALPSSSTSTLLGEDAEVAADFDMEGDDVDDYS-AEVEEL 1025
QY 151 LSO-----SAIQNSKPEGSGFOYPASYNQTLAIGETTPSQLPARGTQA--RATGQSFS 203
Db 1026 LPOHLQPSSSSSGLGTSPSSS--PRT-SPCQSPITISGVPVSLPIRPSRAKSPRTGPPSA 1081
QY 204 QGT---TSRACHLAGEPEA-----PPPPPPPREPPAPSLGSAFHLDPADPAAAAAALYSS 256
Db 1082 QSSPIDAQAPATPLPQKDPQAQLEPKRPPPR-----PVAPP----- 1117
QY 257 STLPAAPPGGSPLAAPQGGSPTKLORGG-SAPEGATYAAPRGSSPKQSPSRLAK---SYS 312
Db 1118 -TRAPAPQRPPEPGASRPATRKKEFGIGAPPSGVARRMEAPK-SPGTTKRDNIGRS 1175
QY 313 TSSPITNIV--SSAGLPIRVTSPTVQSTISSPIHQLSSTIGTYATLSPTKRLVHASE 370
Db 1176 QPSFOAGIAGPGPAGYSTARTIPPRA-GVISAQSHARASA-----GRLTP-----E 1222
QY 371 QYKHSQELVATLQRPGLAGSRASYSQHGHLGPELRAIQSPENHIDPIYED--- 426
Db 1223 SQSTSETSKSGSTFLPEP---LKPQAAPPQ--SLPPPAQRLQEP---LVPAAPMPQS 1274
QY 427 -----RYOKPPMRSLQSOGDPLPPAHTGTYRTSTAPSSPGVDVSVLQRTGSHGPON 480
Db 1275 GPQNLFTPPQPPRRSRSHS---LPSEASSQOVKTINGISDGKRRESPLKIDPPEDLSFN 1331
QY 481 AAAATFORASTAAGPASNVADPYRQLQYCPSESVPYSKSG-----PALPPEGTLARSPSI 535
Db 1332 LLAVSKAQLSVQTSVPPT-PDPKRLIQLPSATQSNVNTLSSVSCMPTMP--IPARSQSQ 1388
QY 536 DSIGKDPREF-----GWRDELPREVIOMLQHOEP 564
Db 1389 ENMRSSNPFTITGLTNTNPFSDRTAAPGNPRAKSESEATSWFSKEEPTVIS---PPF 1444
QY 565 SVQSNAAAYLQHLQFGDNKIKAEIRROG 592
Db 1445 SLOP-----LGHNKGRASSSLDG 1462
```

Search completed: July 19, 2001, 16:06:39
Job time: 199 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2001, 16:06:15 ; Search time 24.53 Seconds

(without alignments)
3804.070 Million cell updates/sec

Title: US-09-501-171A-4

Perfect score: 1225

Sequence: 1 MFARKPPGAAPLGAMPVPO.....PYSELNYESHYPASPDSSW 1225

Scoring table: OLIGO Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size: 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	284	23.2	1247	2 T42209	neural plakophilin
2	11	0.9	911	2 S28498	gene p120 protein
3	11	0.9	1050	2 T26395	hypothetical prote
4	10	0.8	1336	2 T02736	probable SCARECROW
5	9	0.7	153	2 T31654	hypothetical prote
6	9	0.7	184	2 S78091	endocuticular prot
7	9	0.7	243	2 S28444	comb protein - pse
8	9	0.7	301	2 T50001	hypothetical prote
9	9	0.7	415	1 A34170	acrosin (EC 3.4.21
10	9	0.7	448	2 S33926	Wilms' tumor prote
11	9	0.7	449	2 A38080	Wilms' tumor suscep
12	9	0.7	449	2 A39692	Wilms' tumor prote
13	9	0.7	494	2 A42170	zinc finger protei
14	9	0.7	497	2 T27012	hypothetical prote
15	9	0.7	497	2 JC5076	myc-associated zin
16	9	0.7	512	2 S28267	phosphorin I prec
17	9	0.7	543	2 S25128	61k protein - Auto
18	9	0.7	600	2 T14463	galactolactone d
19	9	0.7	647	2 S06450	steroid hormone re
20	9	0.7	658	2 T04831	probable serine/th
21	9	0.7	817	2 S51342	verprolin - yeast
22	9	0.7	860	1 EAMS	elastin precursor
23	9	0.7	921	2 A33718	retinoblastoma pro
24	9	0.7	928	1 RBH9	retinoblastoma-ass
25	9	0.7	967	2 T40171	probable chromosom
26	9	0.7	979	2 T50956	hypothetical prote
27	9	0.7	1333	2 T78557	N-methyl-D-asparta
28	9	0.7	1333	2 S27224	N-methyl-D-asparta
29	9	0.7	1356	1 C45219	N-methyl-D-asparta

30	9	0.7	1420	2 T37781	probable cytoskele
31	9	0.7	1684	2 T02632	hypothetical prote
32	9	0.7	1737	2 T00209	MEG8 protein - hu
33	9	0.7	1953	2 S63244	BN1 protein - yea
34	8	0.7	10	2 A36454	trypsin-modulating
35	8	0.7	15	2 PT0037	light harvesting c
36	8	0.7	70	2 A96650	protein F2401.6 [1
37	8	0.7	74	2 B40513	hypothetical prote
38	8	0.7	92	2 B48831	vitelline membrane
39	8	0.7	102	2 T16978	ribosomal protein
40	8	0.7	105	2 T22564	hypothetical prote
41	8	0.7	117	2 T46322	hypothetical prote
42	8	0.7	121	2 S17718	anther-specific pr
43	8	0.7	121	2 S12245	anther-specific pr
44	8	0.7	135	2 S34815	nifH protein - Rn
45	8	0.7	141	2 A34043	hypothetical proli

ALIGNMENTS

RESULT 1
T42209
neural plakophilin related arm-repeat protein NRPAP - mouse
N:Alternate names: plakoglobin/armadillo protein
C:Species: Mus musculus (house mouse)
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T42209
R:Patfenholz, R.; Franke, W.W.
A:Title: Identification and localization of a neurally expressed member of the plakog
A:Reference number: 222075; MUID:98002299
A:Accession: T42209
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1247 <PAF>
A:Cross-references: EMBL:U90331; NID:92580536; PID:92580537; PIDN:AAB82409.1
A:Experimental source: brain
C:Genetics:
A:Gene: NRPAP
A>Note: predominantly, if not exclusively, expressed in neural and neuroendocrine tis

Query Match 23.2% Score 284; DB 2; Length 1247;
Best Local Similarity 100.0%; Pred. No. 1.5e+260; Indels 0; Gaps 0;
Matches 284; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 514 SPYKSGPALPEGLTARSPSIDSIQKDPREFGWRDPPEVITMLQHOPPSVSNAAAY 573
|||||
DB 511 SPYKSGPALPEGLTARSPSIDSIQKDPREFGWRDPPEVITMLQHOPPSVSNAAAY 570
|||||
OY 574 LQHLFGDNKIKAEIRROGGIOLVDLLDHRTMEVHRSACALRNLYVGKANDNKIALK 633
|||||
DB 571 LQHLFGDNKIKAEIRROGGIOLVDLLDHRTMEVHRSACALRNLYVGKANDNKIALK 630
|||||
OY 634 NCGGIPALVRLRKTTDEIRBELVGVLMNLSGODALMPITODALVLTAAVLIIPSGW 693
|||||
DB 631 NCGGIPALVRLRKTTDEIRBELVGVLMNLSGODALMPITODALVLTAAVLIIPSGW 690
|||||
OY 694 ENSPLQDDRRKIDLHSSQVLRNATGCLRNVSAGGEARRRMECDGIDPALLVYQSLAGS 753
|||||
DB 691 ENSPLQDDRRKIDLHSSQVLRNATGCLRNVSAGGEARRRMECDGIDPALLVYQSLAGS 750
|||||
OY 754 SEIDSKTYENCVCILRNLSYRLAETSGOGHMGTDLDLGLLGE 797
|||||
DB 751 SEIDSKTYENCVCILRNLSYRLAETSGOGHMGTDLDLGLLGE 794
|||||
RESULT 2
S28498
gene p120 protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change 05-Nov-1999

C:Accession: I48701; S28498
 R:Reynolds, A.B.; Herbert, L.; Cleveland, J.L.; Berg, S.T.; Gaut, J.R.
 Oncogene 7, 2439-2445, 1992
 A:Title: p120, a novel substrate of protein tyrosine kinase receptors and of p60v-src, A:Reference number: I48701; MUID:93096477
 A:Accession: I48701
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-911 <RES>
 A:Cross-references: EMBL:Z17804; NID:953544; PIDN:CAV79078.1; PID:953545
 C:Genetics:
 A:Gene: p120
 C:Keywords: cytoskeleton

Query Match 0.9%; Score 11; DB 2; Length 911;
 Best Local Similarity 100.0%; Pred. No. 0.12;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 568 SNAAYLQHLIC 578
 |||||
 DB 384 SNAAYLQHLIC 394

RESULT 3
 T26395
 hypothetical protein Y105C5B.s - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T26395
 R:McMurray, A.
 submitted to the EMBL Data Library, September 1999
 A:Reference number: Z20208
 A:Accession: T26395
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1050 <WIL>
 A:Cross-references: EMBL:AL110479; NID:e1542153; PIDN:CAB54367.1; CESP:Y105C5B.s
 A:Experimental source: clone Y105C5B
 C:Genetics:
 A:Gene: CESP:Y105C5B.s
 A:introns: 40/3; 105/1; 217/3; 278/1; 422/1; 482/1; 850/1; 903/1; 959/2

Query Match 0.9%; Score 11; DB 2; Length 1050;
 Best Local Similarity 100.0%; Pred. No. 0.14;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 908 VVCAYATLRN 918
 |||||
 DB 828 VVCAYATLRN 838

RESULT 4
 T02736
 probable SCARECROW gene regulator [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 02-Feb-2001
 C:Accession: T02736; A84692
 R:Rounsley, S.D.; Lin, X.; Kaul, S.; Shea, T.P.; Fujii, C.Y.; Mason, T.M.; Shen, M.; Rott
 submitted to the EMBL Data Library, August 1998
 A:Description: Arabidopsis thaliana chromosome II BAC T914 genomic sequence.
 A:Reference number: Z14710
 A:Accession: T02736
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1336 <ROU>
 A:Cross-references: EMBL:AC005315; NID:93461834; PID:93461846
 A:Experimental source: cultivar Columbia
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487
 A:Accession: A84692
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1336 <STO>
 A:Cross-references: GB:AE002093; NID:93461846; PIDN:ACG3232.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g29060; T914.14
 A:Map position: 2
 A:introns: 694/2

Query Match 0.8%; Score 10; DB 2; Length 1336;
 Best Local Similarity 100.0%; Pred. No. 1.5;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 810 GKRRKKRSQ 819
 |||||
 DB 942 GKRRKKRSQ 951

RESULT 5
 T31654
 hypothetical protein Y57A10A.gg - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T31654
 R:Smyle, R.
 submitted to the EMBL Data Library, September 1999
 A:Reference number: Z21048
 A:Accession: T31654
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-153 <WIL>
 A:Cross-references: EMBL:AL117195; NID:e1549729; PIDN:CAB55037.1; CESP:Y57A10A.gg
 A:Experimental source: clone Y57A10A
 C:Genetics:
 A:Gene: CESP:Y57A10A.gg

Query Match 0.7%; Score 9; DB 2; Length 153;
 Best Local Similarity 100.0%; Pred. No. 2.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 218 PAPPPPP 226
 |||||
 DB 80 PAPPPPP 88

RESULT 6
 S78091
 endocuticular protein Sgabd-1 - desert locust
 C:Species: Schistocerca gregaria (desert locust)
 C:Date: 04-Dec-1997 #sequence_revision 12-Dec-1997 #text_change 11-Jan-2000
 C:Accession: S78091
 R:Andersen, S.
 submitted to the Protein Sequence Database, November 1997
 A:Description: Amino acid sequence studies on endocuticular proteins from the desert
 A:Reference number: S78091
 A:Accession: S78091
 A:Molecule type: protein
 A:Residues: 1-184 <AND>
 A:Experimental source: strajn albino; adult; abdominal cuticle
 C:Superfamily: cuticle protein LCPI
 C:Keywords: blocked amino end; glycoprotein; pyroglutamic acid
 F:1-184/Product: endocuticular protein Sgabd-1 #status experimental <MAT>
 F:1/Modified site: pyroglutamate carboxylic acid (Gln) #status experimental
 F:24/Binding site: carbonyl group (Thr) (covalent) #status experimental
 F:29,33/Binding site: carbohydrate (Ser) (covalent) #status experimental

Query Match 0.7%; Score 9; DB 2; Length 184;

```

Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 219 APPPPPPR 227
    |||||
Db 35 APPPPPPR 43

RESULT 7
S28444
tonb protein - Pseudomonas putida
C:Species: Pseudomonas putida
C:Date: 07-May-1993 #sequence_revision 07-May-1993 #text_change 24-Nov-1999
C:Accession: S28444
R:Bitter, W.; Tommassen, J.; Weisbeek, P.J.
Mol. Microbiol. 7, 117-130, 1993
A:Title: Identification and characterization of the exbB, exbD and tonB genes of Pseudomonas putida
A:Reference number: S28442; MUID:93172953
A:Accession: S28444
A:Molecule type: DNA
A:Residues: 1-243 <BIT>
A:Cross-references: EMBL:X70139; NID:g49192; PIDN:CAA49716.1; PID:g49195
C:Genetics:
A:Gene: tonB
C:Superfamily: tonb protein

Query Match
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 216 PEPAPPPP 224
    |||||
Db 53 PEPAPPPP 61

RESULT 8
T50001
hypothetical protein F12B17.270 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Sep-2000
C:Accession: T50001
R:Bevan, M.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z25026
A:Accession: T50001
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-301 <BEV>
A:Cross-references: EMBL:AL353995; GSPDB:GN00063; ATSP:F12B17.270
A:Experimental source: cultivar Columbia; BAC clone F12B17
C:Genetics:
A:Gene: ATSP:F12B17.270
A:Map position: 5
C:Superfamily: RING finger homology
F:131-182/Domain: RING finger homology <RRN>

Query Match
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 PPPPPPPR 228
    |||||
Db 29 PPPPPPPR 37

RESULT 9
A34170
acrosin (EC 3.4.21.10) precursor - pig
N:Alternate names: 53K fucose-binding protein
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

```

```

C:Accession: A34170; S08994; S02428; S04940; S16657; S02780; S10695; S12968
R:Baba, T.; Kashiwabara, S.; Watanabe, K.; Itoh, H.; Michikawa, Y.; Kimura, K.; Takad
J. Biol. Chem. 264, 11920-11927, 1989
A:Title: Activation and maturation mechanisms of boar acrosin zymogen based on the de
A:Reference number: A34170; MUID:9308595
A:Accession: A34170
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-415 <BAB>
A:Cross-references: GB:J04950; NID:g164702; PIDN:AAA31131.1; PID:g164703
R:Ceohova, D.; Toepfer-Petersen, E.; Zucker, A.; Jonakova, V.
Biol. Chem. Hoppe-Seyler 371, 317-323, 1990
A:Title: Is sperminogen a modified proacrosin? Isolation, purification, and partial c
A:Reference number: S08994; MUID:90233635
A:Accession: S08994
A:Molecule type: protein
A:Residues: 'X', 18, 'X', 20-25, 'X', 27-32, 'X', 34-38, 'X', 40-50 <CEC>
R:Toepfer-Petersen, E.; Henschen, A.
FEBS Lett. 226, 38-42, 1987
A:Title: Acrosin shows zona and fucose binding, novel properties for a serine protein
A:Reference number: S02428; MUID:88083633
A:Accession: S02428
A:Molecule type: protein
A:Residues: 17-32, 40-55 <TOE>
R:Adham, I.M.; Klemm, U.; Maier, W.M.; Hoyer-Fender, S.; Tsousidou, S.; Engel, W.
Eur. J. Biochem. 182, 563-568, 1989
A:Title: Molecular cloning of proacrosin and analysis of its expression pattern in
A:Reference number: S04940; MUID:89325301
A:Accession: S04940
A:Molecule type: mRNA
A:Residues: 1-7, 9-210, 'Q', 212-216, 'VT', 219-346, 'A', 348-388, 390-393, 'GN', 396, 'LVE', 399
A:Cross-references: EMBL:X14844
A:Note: the authors translated the codon CCG for residue 240 as Ala, GCC for residue
R:Adham, I.M.
submitted to the EMBL data library, March 1989
A:Reference number: S16657
A:Accession: S16657
A:Molecule type: mRNA
A:Residues: 1-7, 9-210, 'Q', 212-216, 'VT', 219-346, 'A', 348-388, 390-398, 'KEUL' <AD2>
A:Cross-references: EMBL:X14844; NID:g1867; PIDN:CAA32948.1; PID:g1868
A:Note: the difference at the carboxyl end is due to a frameshift error
R:Baba, T.; Michikawa, Y.; Kawakura, K.; Arai, Y.
FEBS Lett. 244, 132-136, 1989
A:Title: Activation of boar proacrosin is effected by processing at both N- and C-ter
A:Reference number: S02780; MUID:89171246
A:Accession: S02780
A:Molecule type: protein
A:Residues: 17-69 <BA2>
R:Toepfer-Petersen, E.; Steinberger, M.; von Eschenbach, C.E.; Zucker, A.
FEBS Lett. 265, 51-54, 1990
A:Title: Zona pellucida-binding of boar sperm acrosin is associated with the N-termin
A:Reference number: S10695; MUID:90306316
A:Accession: S10695
A:Molecule type: protein
A:Residues: 40-62 <TO2>
R:Toepfer-Petersen, E.; Calvete, J.; Schaefer, W.; Henschen, A.
FEBS Lett. 275, 139-142, 1990
A:Title: Complete localization of the disulfide bridges and glycosylation sites in bo
A:Reference number: S12968; MUID:91085546
A:Accession: S12968
A:Molecule type: protein
A:Residues: 17-29, 34-66, 68-91, 94-121, 123-166, 171-184, 190-207, 209-216, 219-228, 231-245;
C:Superfamily: acrosin; trypsin homology
C:Keywords: glycoprotein; hydrolase; serine proteinase; sperm
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-415/Product: acrosin #status experimental <MNT>
F:17-39/Product: acrosin light (A) chain #status experimental <LCH>
F:40-415/Product: acrosin heavy (B) chain #status experimental <HCH>
F:300-374/Region: trypsin homology <TRY>
F:19, 208/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:22-153, 26-160, 71-87, 175-244, 207-223, 234-264/Disulfide bonds: #status experimental
F:86, 140, 238/Active site: His, Asp, Ser #status predicted

```

```

Query Match      0.7%; Score 9; DB 1; Length 415;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 218 PAPPPPPP 226
|||||
DB 342 PAPPPPPP 350

RESULT 10
S33926
WILMS' tumor protein WT1 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 10-May-1996
C:Accession: S33926
R:Sharma, P.M.; Yang, X.; Bowman, M.; Roberts, V.; Sukumar, S.
Cancer Res. 52, 6407-6412, 1992
A:Title: Molecular cloning of rat WILMS' tumor complementary DNA and a study of messenger
A:Reference number: S33926; MUID:93046155
A:Accession: S33926
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-448 <SHA>
A:Cross-references: EMBL:X69716
C:Genetics:
A:Gene: WT1
C:Keywords: tumor suppressor

Query Match      0.7%; Score 9; DB 2; Length 448;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 218 PAPPPPPP 226
|||||
DB 58 PAPPPPPP 66

RESULT 11
A38080
WILMS tumor susceptibility protein WT1 - human
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 20-Jun-2000
C:Accession: A38080; S08273; A34673; I38504; I52811; I58315; A54411; S26286
R:Gesliar, M.; Konig, A.; Bruns, G.A.
Genomics 12, 807-813, 1992
A:Title: The genomic organization and expression of the WT1 gene.
A:Reference number: A38080; MUID:92241883
A:Accession: A38080
A:Molecule type: DNA
A:Residues: 1-449 <GSL>
A:Cross-references: GB:X61611; GB:S99414; NID:937981; PIDN:CAA43819.1; PID:9825731
A:Note: It is uncertain whether Met-1 is the initiator or whether translation is initiated
A:Note: sequence extracted from NCBI backbone (NCBIN:99414, NCBIN:99422, NCBIN:99479, NC
A:Note: the sequence in Genbank entry HSWTGE1, release 113.0, PIDN:CAA43819.1 differs
R:Gesliar, M.; Poustra, A.; Cavenee, W.; Neve, R.L.; Orkin, S.H.; Bruns, G.A.P.
Nature 343, 774-778, 1990
A:Title: Homozygous deletion in WILMS tumours of a zinc-finger gene identified by chromo
A:Reference number: S08273; MUID:90158822
A:Accession: S08273
A:Molecule type: mRNA
A:Residues: 580RHPALNPTACRPHRRPSIRPTNHPRACTAAGACPRRLAAILDELIDOPASTCYPERASQHT
A:Cross-references: EMBL:X51630; NID:937977; PIDN:CAA35956.1; PID:937978
R:Call, K.M.; Glaser, T.; Ito, C.Y.; Buckler, A.J.; Pelletier, J.; Haber, D.A.; Rose, E.
Cell 60, 509-520, 1990
A:Title: Isolation and characterization of a zinc finger polypeptide gene at the human c
A:Reference number: A34673; MUID:90150277
A:Accession: A34673
A:Molecule type: mRNA
A:Residues: 85-249, 267-364, 'F', 366-386, 'T', 388-407, 411-449 <CAL>
A:Cross-references: GB:M30393; NID:9340381; PIDN:AAA36810.1; PID:9340382

```

```

R:Haber, D.A.; Sohn, R.L.; Buckler, A.J.; Pelletier, J.; Call, K.M.; Housman, D.E.
Proc. Natl. Acad. Sci. U.S.A. 88, 9618-9622, 1991
A:Title: Alternative splicing and genomic structure of the WILMS tumor gene WT1.
A:Reference number: A56411; MUID:92052142
A:Contents: annotation; alternative splicing
R:Phelan, S.A.; Lindberg, C.; Call, K.M.
Cell Growth Differ. 5, 677-686, 1994
A:Title: WILMS' tumor gene, WT1, mRNA is down-regulated during induction of erythroid
A:Reference number: I38504; MUID:94368704
A:Accession: I38504
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-18 <PHE>
A:Cross-references: EMBL:U06486; NID:9473563; PIDN:AAA62865.1; PID:9458432
R:Pelletier, J.; Brunning, W.; Kashan, C.E.; Mauer, S.M.; Manivel, J.C.; Striegel, J.
Cell 67, 437-447, 1991
A:Title: Germ-line mutations in the WILMS' tumor suppressor gene are associated with a
A:Reference number: I52811; MUID:92005721
A:Accession: I52811
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 355-365, 'H', 367-377 <PEL>
A:Cross-references: GB:S61513; NID:9237599; PIDN:AA520109.1; PID:9237600
A:Note: mutant form
R:Hamilton, T.B.; Barilla, K.C.; Romanik, P.J.
Nucleic Acids Res. 23, 277-284, 1995
A:Title: High affinity binding sites for the WILMS' tumour suppressor protein WT1.
A:Reference number: I58315; MUID:95166649
A:Accession: I58315
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 'MGNHNNHNNHSSGIEGRHW', 301-364, 'F', 366-386, 'T', 388-407, 411-449 <HAN>
A:Cross-references: GB:S75264; NID:9896246; PIDN:AA33443.1; PID:9896247
A:Note: this sequence is engineered
C:Genetics:
A:Gene: GDB:WT1
A:Cross-references: GDB:120496; OMIM:194070
A:Map position: 11p13-11p13
A:Introns: 148/1; 189/1; 223/2; 249/2; 266/2; 298/3; 349/1; 379/1; 410/1
A:Note: mRNA transcripts containing both alternatively spliced regions are the most a
C:Keywords: alternative splicing; DNA binding; kidney; tumor suppressor; zinc finger
F:1-407/Product: WILMS tumor susceptibility protein WT1, splice form 1 #status predic
F:1-407, 411-449/Product: WILMS tumor susceptibility protein WT1, splice form 4 #statu
F:1-249, 267-449/Product: WILMS tumor susceptibility protein WT1, splice form 3 #statu
F:1-249, 267-407, 411-449/Product: WILMS tumor susceptibility protein WT1, splice form

Query Match      0.7%; Score 9; DB 2; Length 449;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 218 PAPPPPPP 226
|||||
DB 58 PAPPPPPP 66

RESULT 12
A39692
WILMS' tumor protein analog, WT1 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 30-Dec-1991 #sequence_revision 30-Dec-1991 #text_change 16-Feb-1997
C:Accession: A39692
R:Buckler, A.J.; Pelletier, J.; Haber, D.A.; Glaser, T.; Housman, D.E.
Mol. Cell. Biol. 11, 1707-1712, 1991
A:Title: Isolation, characterization, and expression of the murine WILMS' tumor gene
A:Reference number: A39692; MUID:91141522
A:Accession: A39692
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-449 <BU>
A:Cross-references: GB:M55512
C:Keywords: alternative splicing; DNA binding; transcription regulation; tumor suppre

```

Query Match 0.7%: Score 9; DB 2; Length 449;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 PAPP PPPP 226
DB 59 PAPP PPPP 67

RESULT 13

A42170

zinc finger protein MAZ - human (fragment)

N:Alternate names: MYC-associated zinc finger protein MAZ; zinc finger protein ZF87

C:Species: Homo sapiens (man)

C>Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 03-Jun-1996

C:Accession: A42170; A46153

R:Pyrc, J.J.; Moberg, K.H.; Hall, D.J.

Biochemistry 31, 4102-4110, 1992

A:Title: Isolation of a novel cDNA encoding a zinc-finger protein that binds to two sites

A:Reference number: A42170; MUID:92232709

A:Accession: A42170

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-494 <PVR>

A:Cross-references: GB:J05371

A:Note: It is uncertain whether Met-18 is the initiator or whether translation is initiated

R:Bosstone, S.A.; Asselin, C.; Patel, A.J.; Marcu, K.B.

Proc. Natl. Acad. Sci. U.S.A. 89, 7452-7456, 1992

A:Title: MAZ, a zinc finger protein, binds to c-MYC and C2 gene sequences regulating tr

A:Reference number: A46153; MUID:92366479

A:Accession: A46153

A:Molecule type: mRNA

A:Residues: 18-417, 'L', A19-494 <BOS>

A:Cross-references: GB:M94046

A:Experimental source: HeLa cells

A:Note: Sequence extracted from NCBI backbone (NCBIN:110666, NCBI:P.110667)

C:Keywords: DNA binding; zinc finger

F:113-125/Region: alanine-rich

F:174-183/Region: alanine-rich

F:207-230/Region: zinc finger

F:296-318/Region: zinc finger

F:324-346/Region: zinc finger

F:354-368/Region: zinc finger

F:373-405/Region: zinc finger

F:409-430/Region: zinc finger

F:452-468/Region: alanine-rich

Query Match 0.7%: Score 9; DB 2; Length 494;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 PAPP PPPP 226
DB 148 PAPP PPPP 156

RESULT 14

T27012

hypothetical protein Y48B1B.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T27012

R:Murray, A.

Submitted to the EMBL Data Library, March 1997

A:Reference number: Z20299

A:Accession: T27012

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-497 <NTL>

A:Cross-references: EMBL:Z93393; PDB: CAB07688.1; GSPDB: GNO0020; CESP: Y48B1B.1

C:Genetics:
A:Gene: CESP:Y48B1B.1
A:Map position: 2
A:Introns: 68/3; 277/1

Query Match 0.7%: Score 9; DB 2; Length 497;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 PAPP PPPP 226
DB 355 PAPP PPPP 363

RESULT 15

JC5076

myc-associated zinc-finger protein - human

N:Alternate names: MAZ protein

C:Species: Homo sapiens (man)

C>Date: 31-Jan-1997 #sequence_revision 31-Jan-1997 #text_change 05-Nov-1999

C:Accession: JC5076

R:Tsutsui, H.; Sakatsune, O.; Itakura, K.; Yokoyama, K.K.

Biochem. Biophys. Res. Commun. 226, 801-809, 1996

A:Title: Members of the MAZ family: A novel cDNA clone for MAZ from human pancreatic

A:Reference number: JC5076; MUID:96428591

A:Accession: JC5076

A:Molecule type: mRNA

A:Residues: 1-497 <TSU>

A:Cross-references: DDBJ:D85131; NID:q1752741; PIDN:BAI12728.1; PID:d1013410; PID:q17

A:Experimental source: pancreatic islet

C:Comment: This protein plays a role in the control of transcriptional initiation of

and between the introns of the mouse gene for immunoglobulin M-D.

C:Keywords: phosphoprotein; zinc finger

F:146,204/480/Binding site: phosphate (Ser) (covalent) (by casein kinase II) #status

F:349/Binding site: phosphate (Tyr) (covalent) #status predicted

Search completed: July 19, 2001, 16:08:37
Job time: 142 sec

Query Match 0.7%: Score 9; DB 2; Length 497;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 PAPP PPPP 226
DB 157 PAPP PPPP 165

Fri Jul 20 10:09:36 2001

us-09-501-171a-4.oli.rpr

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2001, 16:08:11 ; Search time 14.7 Seconds

(without alignments)
2854.624 Million cell updates/sec

Title: US-09-501-171A-4

Perfect score: 1225

Sequence: 1 MFARKPGADPLGAMPVDPQ.....PYSELNVTSHYPASPDSPW 1225

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 93435 seqs, 34255486 residues

Word size: 0

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: SwissProt_39:*

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	11	0.9	911	P120_MOUSE	P30999 mus muscula
2	11	0.9	962	ARVC_HUMAN	000192 homo sapien
3	11	0.9	969	ARVC_MOUSE	P98203 mus muscula
4	9	0.7	243	TONB_PSEPU	Q05613 pseudomonas
5	9	0.7	288	FXE3_MOUSE	Q99744 mus muscula
6	9	0.7	415	ACRO_PIG	P08001 sus scrofa
7	9	0.7	448	WT1_RAT	P49952 rattus norv
8	9	0.7	449	WT1_HUMAN	P19544 homo sapien
9	9	0.7	449	WT1_MOUSE	P22561 mus muscula
10	9	0.7	449	WT1_PIG	P62651 sus scrofa
11	9	0.7	470	SOX8_CHICK	P57074 gallus gall
12	9	0.7	477	MAZ_HUMAN	P56270 homo sapien
13	9	0.7	512	PER1_VOLCA	P81131 volvox cart
14	9	0.7	520	WASP_MOUSE	P70315 mus muscula
15	9	0.7	543	VP61_NPVAC	Q03209 autographa
16	9	0.7	621	VP40_HSVBC	P54817 bovine herp
17	9	0.7	647	KNRL_DROME	P13054 drosophila
18	9	0.7	817	VRP1_YEAST	P37370 saccharomyc
19	9	0.7	837	SM4G_MOUSE	Q99747 mus muscula
20	9	0.7	839	APB1_RAT	P54305 mus muscula
21	9	0.7	860	ELS_MOUSE	P13405 mus muscula
22	9	0.7	921	RB_MOUSE	P06400 homo sapien
23	9	0.7	928	RB_HUMAN	Q00391 mus muscula
24	9	0.7	1323	NME4_MOUSE	Q03391 mus muscula
25	9	0.7	1353	BN11_YEAST	P41832 saccharomyc
26	9	0.7	1953	TRX2_HUMAN	Q99747 mus muscula
27	9	0.7	2715	TMOE_AEDAE	P19425 aedes aegypt
28	8	0.7	121	ASF2_HELAN	P22184 helianthus
29	8	0.7	135	NTU1_RHOCA	Q07178 rhodobacter
30	8	0.7	141	YPRO_OMEFU	P21260 owenla fusi
31	8	0.7	142	YMB8_YEAST	Q03525 saccharomyc
32	8	0.7	144	YHP5_YEAST	P18808 saccharomyc
33	8	0.7	160	EZRI_RAT	P31977 rattus norv

34	8	0.7	174	1	THBB_RAT	P17491 rattus norv
35	8	0.7	182	1	SP23_TENMO	Q27022 xenopus mo
36	8	0.7	196	1	RBMS_XENLA	Q97955 xenopus lae
37	8	0.7	215	1	HANI_HUMAN	Q96004 homo sapien
38	8	0.7	223	1	TONB_XANCP	Q34261 xanthomonas
39	8	0.7	231	1	CUT2_CAEEL	P34682 caenorhabdi
40	8	0.7	261	1	GSH1_MOUSE	P31335 mus muscula
41	8	0.7	271	1	DB83_CAEEL	Q93470 caenorhabdi
42	8	0.7	297	1	HX1M_CHICK	Q93367 gallus gall
43	8	0.7	299	1	NMS3_MYCTU	Q10390 mycobacteri
44	8	0.7	309	1	HXA4_CHICK	P17277 gallus gall
45	8	0.7	310	1	IF2B_SCHHO	P56329 schizosacch

ALIGNMENTS

RESULT	ID	STANDARD	PRT	911 AA.
1	P120_MOUSE			
P120_MOUSE				
AC	P30999;			
DT	01-JUL-1993 (Rel. 26, Created)			
DT	01-JUL-1993 (Rel. 26, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	P120 PROTEIN.			
GN	P120 OR CATNS.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=SWISS;			
RC	MEDLINE=93096477; PubMed=1334250;			
RA	Reynolds A.B., Herbert L., Cleveland J.L., Berg S.T., Gant J.R.;			
RT	"p120, a novel substrate of protein tyrosine kinase receptors and of			
RT	p60v-src, is related to cadherin-binding factors beta-catenin,			
RT	Plakoglobin and armadillo."			
RL	Oncogene 7:2439-2445(1992).			
RN	[2]			
RP	REVISIONS.			
RA	Reynolds A.B.;			
RL	Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.			
CC	-1- P120 PHOSPHORYLATED BY PROTEIN-TYROSINE KINASES.			
CC	-1- SIMILARITY: BELONGS TO THE BETA-CATENIN FAMILY.			
CC	-1- SIMILARITY: CONTAINS 4 ARM REPEATS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
CC	-----			
DR	EMBL: Z17804; CAA79078.1; -			
DR	PIR: S28498; S28498.			
DR	MGI: MGI:105100; Catns.			
DR	InterPro: IPR000225; -			
DR	Pfam: PF00514; Armadillo_seg. 4.			
DR	PROSITE: PS50176; ARM_REPEAT; 3.			
KW	Cytoskeleton; Structural protein; Phosphorylation; Repeat.			
FT	REPEAT 398..437 ARM 1.			
FT	REPEAT 441..461 ARM 2.			
FT	REPEAT 647..687 ARM 3.			
FT	REPEAT 694..733 ARM 4.			
FT	SEQUENCE 911 AA; 101731 MW; EEL18C623948DD1 CRC64;			

Query Match 0.98; Score 11; DB 1; Length 911;
Best Local Similarity 100.0%; Pred. NO. 0.057;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 568 NNAAYLQHLC 578
 DB 384 NNAAYLQHLC 394

RESULT 2
 ARVC_HUMAN STANDARD; PRT; 962 AA.
 AC 000192;
 DT 01-OCT-2000 (Rel. 40, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE ARADILLO REPEAT PROTEIN DELETED IN VELO-CARDIO-FACIAL SYNDROME.
 GN ARVCF.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NX NCBL_TaxID=9606;
 RA [1]
 RP SEQUENCE FROM N.A. (SHORT AND LONG ISOFORMS).
 RX MEDLINE=97271559; PubMed=9126485;
 RA Strotkin H., O'Donnell H., Dasgupta R., Halford S., St Jore B.,
 RA Puech A., Parimoo S., Morrow B., Skoultsi A., Weissman S.,
 RA Scambler P., Kuchelapati R.;
 RT "Identification of a new human catenin gene family member (ARVCF) from
 RT the region deleted in velo-cardio-facial syndrome.";
 RL Genomics 41:75-83(1997).
 CC -1- FUNCTION: INVOLVED IN PROTEIN-PROTEIN INTERACTIONS AT ADHERENS
 CC JUNCTIONS.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A SHORT FORM AND A LONG FORM
 CC (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: FOUND IN ALL THE EXAMINED TISSUES INCLUDING
 CC HEART BRAIN, LIVER AND KIDNEY. FOUND AT LOW LEVEL IN LUNG.
 CC -1- DISEASE: GENE DELETED IN VELO-CARDIO-FACIAL SYNDROME (VCFS) AND IS
 CC HEMIZYGOSITY IN ALL VCFS PATIENTS WITH INTERSTITIAL DELETIONS. THIS
 CC PHENOTYPE MAY PLAY A ROLE IN THE ETIOLOGY OF SOME OF THE
 CC PHENOTYPES ASSOCIATED WITH VCFS CHARACTERIZED BY A WIDE SPECTRUM
 CC PHENOTYPES, INCLUDING CONOTRONCAL HEART DEFECTS, CLEFT PALATE AND
 CC FACIAL DYSMORPHOLOGY.
 CC -1- SIMILARITY: BELONGS TO THE BETA-CATENIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 10 ARM REPEATS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U51269; AAC51202.1; --
 DR HSP: Q02248; 2BCT.
 DR MIM: 602269; --
 DR MIM: 192430; --
 DR InterPro: IPR000225; --
 DR Pfam: PF00514; Armadillo_seg. 4.
 DR PROSITE: PS50176; ARM_REPEAT; 3.
 KW Repeat; Coiled coil; Alternative splicing.
 FT DOMAIN 8 46
 FT DOMAIN 607 623 COILED COIL (POTENTIAL).
 FT DOMAIN 608 611 NOCLEAR LOCALIZATION (POTENTIAL).
 FT REPEAT 348 387 POLY-ARG.
 FT REPEAT 390 429 ARM 1.
 FT REPEAT 433 473 ARM 2.
 FT REPEAT 468 508 ARM 3.
 FT REPEAT 526 565 ARM 4.
 FT REPEAT 575 622 ARM 5.
 FT REPEAT 646 686 ARM 6.
 FT REPEAT 699 738 ARM 7.
 FT REPEAT 739 781 ARM 8.
 FT REPEAT 782 826 ARM 9.
 FT REPEAT 826 869 ARM 10.
 FT VARSPLIC 1
 EMBL: NNAAYLQHLC 579
 DB 372 NNAAYLQHLC 382
 EMBL: NNAAYLQHLC 382

FT SO SEQUENCE 962 AA; 104641 MW; 7A1814A022FF2B1 CRC64;
 (IN SHORT ISOFORM).

Query Match 0.9%; Score 11; DB 1; Length 962;
 Best Local Similarity 100.0%; Pred. No. 0.06;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 569 NNAAYLQHLCF 579
 DB 377 NNAAYLQHLCF 387

RESULT 3

ARVC_MOUSE STANDARD; PRT; 969 AA.

AC P98203;
 DT 01-OCT-2000 (Rel. 40, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE ARADILLO REPEAT PROTEIN DELETED IN VELO-CARDIO-FACIAL SYNDROME
 GN ARVCF.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NX NCBL_TaxID=10090;
 RA [1]
 RP SEQUENCE FROM N.A.
 RA Starzinski-Powitz A., Kaufmann U., Urbich C., Zuplinger C.;
 RT "The armadillo repeat protein ARVCF, a candidate for the velo cardio
 RT facial syndrome forms complexes with cadherin.";
 RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: INVOLVED IN PROTEIN-PROTEIN INTERACTIONS AT ADHERENS
 CC JUNCTIONS (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE BETA-CATENIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 10 ARM REPEATS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AJ23418; -- NOT_ANNOTATED_CDS.
 DR MGD: MGI:109620; Arvcf
 DR PROSITE: PS50176; ARM_REPEAT; 3.
 KW Repeat; Coiled coil.
 FT NON_TER 1 1
 FT DOMAIN 328 376 COILED COIL (POTENTIAL).
 FT REPEAT 377 418 ARM 1.
 FT REPEAT 419 462 ARM 2.
 FT REPEAT 463 503 ARM 3.
 FT REPEAT 521 560 ARM 4.
 FT REPEAT 570 616 ARM 5.
 FT REPEAT 634 674 ARM 6.
 FT REPEAT 675 720 ARM 7.
 FT REPEAT 721 769 ARM 8.
 FT REPEAT 770 814 ARM 9.
 FT REPEAT 814 869 ARM 10.
 FT SEQUENCE 969 AA; 105378 MW; 91130069E484C5B5 CRC64;

Query Match 0.9%; Score 11; DB 1; Length 969;
 Best Local Similarity 100.0%; Pred. No. 0.061;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 569 NNAAYLQHLCF 579
 DB 372 NNAAYLQHLCF 382

EMBL: NNAAYLQHLCF 579
 DB 372 NNAAYLQHLCF 382
 EMBL: NNAAYLQHLCF 382

```

RESULT 4
TONB_PSEPU STANDARD; PRT: 243 AA.
ID TONB_PSEPU
AC 005613;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE TONB PROTEIN.
GN TONB.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WCS358;
RX MEDLINE=93172953; Pubmed=8437515;
RA Bilter W., Tommassen J., Weisbeek P.J.;
RT "Identification and characterization of the exbB, exbD and tonB genes
RT of Pseudomonas putida WCS358: their involvement in
RT ferric-pseudobactin transport."
RL Mol. Microbiol. 7:117-130(1993).
CC -1- FUNCTION: INTERACTS WITH OUTER MEMBRANE RECEPTOR PROTEINS THAT
CC CARRY OUT HIGH-AFFINITY BINDING AND ENERGY DEPENDENT UPTAKE INTO
CC THE PERIPLASMIC SPACE OF SPECIFIC SUBSTRATES. IT COULD ACT TO
CC TRANSDUCE ENERGY FROM THE CYTOPLASMIC MEMBRANE TO SPECIFIC ENERGY-
CC REQUIRING PROCESSES IN THE OUTER MEMBRANE, RESULTING IN THE
CC RELEASE INTO THE PERIPLASM OF LIGANDS BOUND BY THESE OUTER
CC MEMBRANE PROTEINS (BY SIMILARITY).
CC -1- SUBUNIT: THE ACCESSORY PROTEINS EXBBD AND EXBD SEEM TO FORM A
CC COMPLEX WITH TONB.
CC -1- SUBCELLULAR LOCATION: PERIPLASMIC. ANCHORED TO THE CYTOPLASMIC
CC MEMBRANE VIA ITS N-TERMINAL SIGNAL-LIKE SEQUENCE, SPANS THE
CC PERIPLASM.
CC -1- SIMILARITY: BELONGS TO THE TONB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X70139; CAA49716.1; -
DR PIR: S28444; S28444.
KM Transport; Protein transport; Inner membrane; Periplasmic;
KW Transmembrane; Signal-anchor; Repeat.
FT DOMAIN 1 12 CYTOPLASMIC (POTENTIAL).
FT TRANSLEM 13 32 SIGNAL-ANCHOR (POTENTIAL).
FT DOMAIN 33 243 PERIPLASMIC (POTENTIAL).
FT DOMAIN 50 61 6 X 2 AA TANDEM REPEATS OF X-P.
SQ SEQUENCE 243 AA; 25993 MW; 6575C15A147A2774 CRC64;

Query Match 0.7%; Score 9; DB 1; Length 243;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 216 PEPAPPPP 224
DB 53 PEPAPPPP 61

```

```

GN FOXE3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV; TISSUE=Lens;
RX MEDLINE=20119184; Pubmed=10652278;
RA Blixt A., Mahlapuu M., Altolu M., Peltto-Huikko M., Eerback S.,
RA Carlsson P.;
RT "A forkhead gene, foxe3, is essential for lens epithelial
RT proliferation and closure of the lens vesicle."
RL Genes Dev. 14:245-254(2000).
CC -1- FUNCTION: TRANSCRIPTION FACTOR ESSENTIAL FOR LENS EPITHELIAL
CC PROLIFERATION AND CLOSURE OF THE LENS VESICLE.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN THE DEVELOPING LENS FROM THE
CC START OF LENS PLACODE INDUCTION AND BECOMES RESTRICTED TO THE
CC ANTERIOR PROLIFERATING CELLS WHEN LENS FIBER DIFFERENTIATION
CC BEGINS.
CC -1- DISEASE: DEFECTS IN FOXE3 IS A CAUSE OF DYSGENETIC LENS (DYL); IN
CC MOUSE MUTANT DYL THE LENS VESICLE FAILS TO SEPARATE FROM THE
CC ECTODERM, CAUSING A FUSION BETWEEN THE LENS AND THE CORNEA. LACK
CC OF A PROLIFERATING ANTERIOR LENS EPITHELIUM LEADS TO ABSENCE OF
CC SECONDARY LENS FIBERS AND A DYSPLASTIC, CATARACTIC LENS.
CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF142647; AAF15997.1; -
DR MGI: 1353569; Foxe3.
DR InterPro: IPR001766; -
DR Pfam: PF00250; Fork head; 1.
DR PRINTS: PR00053; FORKHEAD.
DR PROSITE: PS00657; FORK_HEAD_1; 1.
DR PROSITE: PS00658; FORK_HEAD_2; 1.
DR PROSITE: PS50039; FORK_HEAD_3; 1.
KW Transcription regulation; DNA-binding; Nuclear protein.
FT DOMAIN 55 58 POLY-ARG.
FT DNA_BIND 63 154 FORK-HEAD.
FT DOMAIN 166 176 POLY-PRO.
FT DOMAIN 179 186 POLY-PRO.
FT DOMAIN 229 232 POLY-ALA.
SQ SEQUENCE 288 AA; 30541 MW; 16425926F3E466C5 CRC64;

Query Match 0.7%; Score 9; DB 1; Length 288;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 PAPPPIPP 226
DB 166 PAPPPIPP 174

```

```

RESULT 5
FOX3_MOUSE STANDARD; PRT: 288 AA.
ID FOX3_MOUSE
AC Q90Y14;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE FORKHEAD BOX PROTEIN E3.

```

```

RESULT 6
ACRO_PIG STANDARD; PRT: 415 AA.
ID ACRO_PIG
AC P08001; P08000;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE ACROSIN PRECURSOR (EC 3.4.21.10) (53 KDA FUCOSE-BINDING PROTEIN).
GN ACR.
OS Sus scrofa (pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89308595; PubMed=2745422;
 RA Bada T., Kashiwabara S.I., Watanabe K., Itoh H., Michikawa Y.,
 RA Kimura K., Takada M., Fukamizu A., Arai Y.;
 RT "Activation and maturation mechanisms of boar acrosin zymogen based
 RT on the deduced primary structure";
 RL J. Biol. Chem. 264:11920-11927(1989).
 RN [12]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Testis;
 RX MEDLINE=89325301; PubMed=2502391;
 RA Adam I.M., Maier W.-M., Hoyer-Fender S., Tsaousidou S., Engel W.,
 RA Klemm U.;
 RT "Molecular cloning of preproacrosin and analysis of its expression
 RT pattern in spermatogenesis";
 RL Eur. J. Biochem. 182:563-568(1989).
 RN [13]
 RP SEQUENCE OF 17-39.
 RC TISSUE-Sperm;
 RX MEDLINE=84261484; PubMed=6378631;
 RA Fock-Nwezel R., Lottspeich F., Henschen A., Mueller-Esterl W.;
 RT "Boar acrosin is a two-chain molecule. Isolation and primary
 RT structure of the light chain; homology with the pro-part of other
 RT serine proteinases";
 RL Eur. J. Biochem. 141:441-446(1984).
 RN [14]
 RP SEQUENCE OF 25-91.
 RC TISSUE-Sperm;
 RX MEDLINE=81115822; PubMed=7007202;
 RA Fock-Nwezel R., Lottspeich F., Henschen A., Mueller-Esterl W.,
 RA Fritz H.;
 RT "N-terminal amino acid sequence of boar sperm acrosin. Homology with
 RT other serine proteinases";
 RL Hoppe-Seyler's Z. Physiol. Chem. 361:1823-1828(1980).
 RN [15]
 RP SEQUENCE OF 17-32 AND 40-53.
 RC TISSUE-Sperm;
 RX MEDLINE=88083633; PubMed=3480243;
 RA Toepfer-Petersen E., Henschen A.;
 RT "Acrosin shows zona and fucose binding, novel properties for a serine
 RT proteinase";
 RL FEBS Lett. 226:38-42(1987).
 RN [16]
 RP SEQUENCE OF 17-40.
 RC MEDLINE=90253655; PubMed=2111146;
 RA Cechova D., Toepfer-Petersen E., Zuckner A., Jonakova V.;
 RT "Is spermogen a modified proacrosin? Isolation, purification, and
 RT partial characterization of low-molecular-mass boar proacrosin";
 RL Biol. Chem. Hoppe-Seyler 371:317-323(1990).
 RN [17]
 RP DISULFIDE BONDS, CARBOHYDRATE-LINKAGE SITES, AND PARTIAL SEQUENCE.
 RX MEDLINE=91083546; PubMed=2261983;
 RA Toepfer-Petersen E., Calvete J.J., Schaefer W., Henschen A.;
 RT "Complete localization of the disulfide bridges and glycosylation
 RT sites in boar sperm acrosin";
 RL FEBS Lett. 275:139-142(1990).
 CC -1- FUNCTION: ACROSIN IS THE MAJOR PROTEASE OF MAMMALIAN SPERMATOZOA.
 CC IT IS A SERINE PROTEASE OF TRYPsin-LIKE CLEAVAGE SPECIFICITY, IT
 CC IS SYNTHESIZED IN A ZYMOGEN FORM, PROACROSIN AND STORED IN THE
 CC ACROSOME.
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF ARG- AND LYS-BONDS; PREFERENTIAL
 CC CLEAVAGE ARG-XAA-> LYS-LYS-> LYS-XAA.
 CC -1- SUBUNIT: HEAVY CHAIN (CATALYTIC) AND A LIGHT CHAIN LINKED BY TWO
 CC DISULFIDE BONDS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPsin FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: J04950; AAA31131.1; -
 CC EMBL: X14844; CAA32948.1; -
 CC PIR: A34170; A34170.
 CC PIR: S02428; S02428.
 CC PIR: S04940; S04940.
 CC PIR: S08994; S08994.
 CC PIR: S12968; S12968.
 CC MEROPS: S01.223; -
 CC InterPro: IPR001254; -
 CC InterPro: IPR001314; -
 CC Pfam: PF00089; trypsin; 1.
 CC PRINTS: PR00722; CHYMOTRYPSIN.
 CC PROSITE: PS00134; TRYPsin_HIS; 1.
 CC PROSITE: PS00135; TRYPsin_SER; 1.
 CC KMW Hydrolyase; Serine protease; Glycoprotein; Zymogen; Sperm; signal.
 FT SIGNAL 1 16
 FT CHAIN 17 415 ACROSIN.
 FT CHAIN 17 39 ACROSIN LIGHT CHAIN.
 FT CHAIN 40 338 ACROSIN HEAVY CHAIN.
 FT PROPEP 339 415 PRO-RICH.
 FT DISULFID 22 152 INTERCHAIN.
 FT DISULFID 26 160 INTERCHAIN.
 FT DISULFID 71 87 INTERCHAIN.
 FT DISULFID 175 244
 FT DISULFID 207 223
 FT DISULFID 234 264
 FT CARBOHYD 19 19 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 208 208 N-LINKED (GLCNAC. . .).
 FT ACT_SITE 86 86 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 140 140 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 238 238 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CONFLICT 8 8 MISSING (IN REF. 2).
 FT CONFLICT 211 211 R -> Q (IN REF. 2).
 FT CONFLICT 217 218 IR -> VT (IN REF. 2).
 FT CONFLICT 347 347 P -> A (IN REF. 2).
 FT CONFLICT 389 389 MISSING (IN REF. 2).
 FT CONFLICT 399 402 RSTY -> KEL (IN REF. 2).
 FT CONFLICT 403 415 MISSING (IN REF. 2).
 SQ SEQUENCE 415 AA; 45387 MW; 5AD1190087E95772 CRC64;
 Query Match 0.7%; Score 9; DB 1; Length 415;
 Best Local Similarity 100.0%; Pred. No. 2.2;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 218 PAPPPEPP 226
 ID PAPPPEPP 350
 Db 342 PAPPPEPP 350
 RESULT 7
 ID WT1_RAT STANDARD; PRT; 448 AA.
 AC P49952;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE WILMS' TUMOR PROTEIN HOMOLOG.
 GN WT1 OR WT-1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Kidney;
 RX MEDLINE=93046155; PubMed=1330293;
 RA Sharma P.M., Yang X., Bowman M., Roberts V., Sukumar S.;

RT "Molecular cloning of rat Wilms' tumor complementary DNA and a study
 of messenger RNA expression in the urogenital system and the brain."
 RL Cancer Res. 52:6407-6412(1992).
 CC - FUNCTION: POTENTIAL ROLE IN TRANSCRIPTIONAL REGULATION. RECOGNIZES
 CC AND BINDS TO THE DNA SEQUENCE 5'-CGCCCCCGC-3'.
 CC - SUBCELLULAR LOCATION: NUCLEAR.
 CC - ALTERNATIVE PRODUCTS: TWO ALTERNATIVE SPLICING SITES EXISTS.
 CC - TISSUE SPECIFICITY: KIDNEY (BY SIMILARITY).
 CC - DEVELOPMENTAL STAGE: EXPRESSED DURING KIDNEY DEVELOPMENT.
 CC - SIMILARITY: BELONGS TO THE EGR FAMILY OF C2H2-TYPE ZINC-FINGER
 CC PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@isb-sdb.ch).
 CC -----
 CC EMBL: X69716; CAA49373.1; -
 CC TRANSFAC: T02352; -
 CC InterPro: IPR000822; -
 CC InterPro: IPR000976; -
 CC Pfam: PF00096; Zf-C2H2; 4.
 CC PRINTS: PR00046; ZINC-FINGER.
 CC PRINTS: PR00049; WILMTUMOR.
 CC PROSITE: PS00028; ZINC_FINGER_C2H2_1; 4.
 CC PROSITE: PS0157; ZINC_FINGER_C2H2_2; 4.
 CC Trunc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;
 CC Transcription regulation; Alternative splicing; Anti-oncogene.
 KW DOMAIN 27
 FT 322 437 ZINC_FINGERS.
 FT ZN_FING 322 346 C2H2-TYPE.
 FT ZN_FING 352 376 C2H2-TYPE.
 FT ZN_FING 382 404 C2H2-TYPE.
 FT ZN_FING 413 437 C2H2-TYPE.
 FT VARSPLIC 249 265 MISSING (IN ISOFORM 2 AND ISOFORM 4).
 FT VARSPLIC 407 409 MISSING (IN ISOFORM 2 AND ISOFORM 4).
 SQ SEQUENCE 448 AA; 49193 MW; 329AC9AC1FF73F76 CRC64;
 Query Match 0.7%; Score 9; DB 1; Length 448;
 Best Local Similarity 100.0%; Pred. No. 2.4;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 218 PAPPPEPP 226
 DB 58 PAPPPEPP 66
 RESULT 8
 WT1_HUMAN STANDARD: PRT; 449 AA.
 ID WT1_HUMAN
 AC P19544;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE WILMS' TUMOR PROTEIN (WT33).
 GN WT1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Fetal kidney;
 RX MEDLINE=90158822; PubMed=2154702;
 RA Gessler W., Poustka A., Cavenee W., Neve R.L., Orkin S.H.,
 RA Bruns G.A.P.;
 RT "Homozygous deletion in Wilms tumors of a zinc-finger gene
 RT identified by chromosome jumping.";
 RL Nature 343:774-778(1990).

RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Placenta;
 RX MEDLINE=92052142; PubMed=1658787;
 RA Haber D.A., Sohn R.L., Buckler A.J., Pelletier J., Call K.M.,
 RA Housman D.E.;
 RT "Alternative splicing and genomic structure of the Wilms tumor gene
 RT WT1.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:9618-9622(1991).
 RN [3]
 RP SEQUENCE OF 85-449 FROM N.A.
 RX MEDLINE=90150277; PubMed=2154335;
 RA Call K.M., Glaser T., Ito C.Y., Buckler A.J., Pelletier J.,
 RA Haber D.A., Rose E.A., Kral A., Yeager H., Lewis W.H., Jones C.,
 RA Housman D.E.;
 RT "Isolation and characterization of a zinc finger polypeptide gene at
 RT the human chromosome 11 Wilms' tumor locus.";
 RL Cell 60:509-520(1990).
 RN [4]
 RP IDENTIFICATION OF START CODON AND ALTERNATIVE SPLICING SITES.
 RX MEDLINE=91141522; PubMed=1671709;
 RA Buckler A.J., Pelletier J., Haber D.A., Glaser T., Housman D.E.;
 RT "Isolation, characterization, and expression of the murine Wilms'
 RT tumor gene (WT1) during kidney development.";
 RL Mol. Cell. Biol. 11:1707-1712(1991).
 RN [5]
 RP VARIANT WT CYS-366.
 RX MEDLINE=92279213; PubMed=1317572;
 RA Little M.H., Prosser J., Conde A., Smith P.J., van Heyningen V.,
 RA Hastie N.D.;
 RT "Zinc finger point mutations within the WT1 gene in Wilms tumor
 RT patients.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:4791-4795(1992).
 RN [6]
 RP VARIANTS DDS.
 RX MEDLINE=92005721; PubMed=1655284;
 RA Pelletier J., Bruening W., Kashan C.E., Mauer S.M., Manivel J.C.,
 RA Striegel J.E., Houghton D.C., Junien C., Habib R., Fouser L.,
 RA Fine R.N., Silverman B.L., Haber D.A., Housman D.;
 RT "Germline mutations in the Wilms' tumor suppressor gene are
 RT associated with abnormal urogenital development in Denys-Drash
 RT syndrome.";
 RL Cell 67:437-447(1991).
 RN [7]
 RP VARIANTS DDS.
 RX MEDLINE=93265053; PubMed=1338906;
 RA Baird P.N., Santos A., Groves N., Jadresic L., Cowell J.K.;
 RT "Constitutional mutations in the WT1 gene in patients with
 RT Denys-Drash syndrome.";
 RL Hum. Mol. Genet. 1:301-305(1992).
 RN [8]
 RP VARIANTS DDS.
 RX MEDLINE=93271983; PubMed=8388765;
 RA Little M.H., Williamson K.A., Mannens M., Kelsey A., Gosden C.,
 RA Hastie N., van Heyningen V.;
 RT "Evidence that WT1 mutations in Denys-Drash syndrome patients may act
 RT in a dominant-negative fashion.";
 RL Hum. Mol. Genet. 2:259-264(1993).
 RN [9]
 RP VARIANT MESOTHELIOMA GUY-273.
 RX MEDLINE=94004972; PubMed=8401592;
 RA Park S., Schalling M., Bernard A., Maheswaran S., Shipley G.C.,
 RA Roberts D., Fletcher J., Shipman R., Rheinwald J., Demetri G.,
 RA Griffin J., Minden M., Housman D.E., Haber D.A.;
 RT "The Wilms tumor gene WT1 is expressed in murine mesoderm-derived
 RT tissues and mutated in a human mesothelioma.";
 RL Nat. Genet. 4:415-420(1993).
 RN [10]
 RP VARIANTS WT SER-181 AND ALA-253.
 RX MEDLINE=97268681; PubMed=9108089;
 RA Schumacher V., Schneider S., Figue A., Wildhardt G., Harms D.,
 RA Schmidt D., Weirich A., Ludwig R., Royer-Pokora B., Harms D.,
 RT "Correlation of germ-line mutations and two-hit inactivation of the

RT WT1 gene with Wilms tumors of stromal-predominant histology.";
 RL Proc. Natl. Acad. Sci. U.S.A. 94:3972-3977(1997).
 [11]
 RP VARIANTS DMS TYR-377; LEU-383 AND ASN-396.
 RA MEDLINE=98196341; PubMed=9529364;
 RA Jeanpierre C., Denamur E., Henry I., Cabanis M.-O., Luce S.,
 RA Cecille A., Elion J., Peuchmaur M., Lolrat C., Naudet P.,
 RA Gubler M.-C., Junien C.;
 RT "Identification of constitutional WT1 mutations, in patients with
 RT isolated diffuse mesangial sclerosis, and analysis of
 RT genotype/phenotype correlations by use of a computerized mutation
 RT database.";
 RL Am. J. Hum. Genet. 62:824-833(1998).
 [12]
 RP REVIEW.
 RA MEDLINE=92207913; PubMed=1313285;
 RA Haber D.A., Buckler A.J.;
 RT "WT1: a novel tumor suppressor gene inactivated in Wilms' tumor";
 RL New Biol. 4:97-106(1992).
 [13]
 RP REVIEW.
 RA MEDLINE=93345769; PubMed=8393820;
 RA Rauscher F.J. III;
 RT "The WT1 Wilms tumor gene product: a developmentally regulated
 RT transcription factor in the kidney that functions as a tumor
 RT suppressor.";
 RL FASEB J. 7:896-903(1993).
 CC -1- FUNCTION: POTENTIAL ROLE IN TRANSCRIPTIONAL REGULATION. RECOGNIZES
 CC AND BINDS TO THE DNA SEQUENCE 5'-CGCCCCCGC-3'.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE KIDNEY AND A SUBSET OF
 CC HEMATOPOIETIC CELLS.
 CC -1- DISEASE: WILMS TUMOR (WT) IS AN EMBRYONAL MALIGNANCY OF THE
 CC KIDNEY THAT AFFECTS APPROXIMATELY 1 IN 10,000 INFANTS AND YOUNG
 CC CHILDREN. IT OCCURS BOTH IN SPORADIC AND HEREDITARY FORMS.
 CC INACTIVATION OF WT1 IS ONE OF THE CAUSES OF WILMS TUMOR.
 CC -1- DISEASE: PATIENTS WITH DENIS-DRASH SYNDROME (DDS), WHICH IS
 CC CHARACTERIZED BY A TYPICAL NEPHROPATHY AND GENITAL ABNORMALITIES,
 CC HAVE DEFECTS IN THE WT1 GENE.
 CC -1- DISEASE: DEFECTS IN WT1 ARE ALSO A CAUSE OF DIFFUSE MESANGIAL
 CC SCLEROSIS (DMS). A FORM A DDS.
 CC -1- SIMILARITY: BELONGS TO THE EGR FAMILY OF C2H2-TYPE ZINC-FINGER
 CC PROTEINS.
 CC -1- DATABASE: NAME=WT1; NOTE=WT1 mutation database;
 CC WWW="http://www.umd.necker.fr:2003/".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL X51630; CAA35956.1; AL1; INIT.
 DR EMBL M80232; AAA61299.1; JOINED.
 DR EMBL M80217; AAA61299.1; JOINED.
 DR EMBL M80218; AAA61299.1; JOINED.
 DR EMBL M80219; AAA61299.1; JOINED.
 DR EMBL M80220; AAA61299.1; JOINED.
 DR EMBL M80221; AAA61299.1; JOINED.
 DR EMBL M80228; AAA61299.1; JOINED.
 DR EMBL M80229; AAA61299.1; JOINED.
 DR EMBL M80231; AAA61299.1; JOINED.
 DR EMBL M30383; AAA36810.1; JOINED.
 DR PIR A34673; A34673.
 DR PIR S08273; S08273.
 DR TRNSPAC T00899; -.
 DR MIM 194070; -.
 DR MIM 256370; -.
 DR InterPro: IPR000822; -.
 DR

DR InterPro: IPR000976; -.
 DR Pfam: PF00096; zf-C2H2; 4.
 DR PRINTS: PR00048; ZINC_FINGER.
 DR PRINTS: PR00049; WILMSTUMOR.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2.1; 4.
 DR PROSITE: PS0157; ZINC_FINGER_C2H2.2; 4.
 DR Zinc-finger; Metal-binding; Repeat; Nuclear protein;
 KW Transcription regulation; Alternative splicing; Anti-oncogene;
 KW Disease mutation.
 FT DOMAIN 27 83 PRO-RICH.
 FT DOMAIN 323 438 ZINC_FINGERS.
 FT ZN_FING 323 347 C2H2-TYPE.
 FT ZN_FING 353 377 C2H2-TYPE.
 FT ZN_FING 383 405 C2H2-TYPE.
 FT ZN_FING 414 438 C2H2-TYPE.
 FT VARSPIC 250 266 MISSING (IN ISOFORM 2 AND ISOFORM 3).
 FT VARSPIC 408 410 MISSING (IN ISOFORM 2 AND ISOFORM 4).
 FT VARIANT 181 181 P -> S (IN WT).
 FT VARIANT 181 181 /FTID-VAR_007739.
 FT VARIANT 223 223 S -> N (IN WT).
 FT VARIANT 223 223 /FTID-VAR_007740.
 FT VARIANT 253 253 G -> A (IN WT).
 FT VARIANT 253 253 /FTID-VAR_007741.
 FT VARIANT 273 273 S -> G (IN MESOTHELIOMA).
 FT VARIANT 273 273 /FTID-VAR_007742.
 FT VARIANT 330 330 C -> Y (IN DDS).
 FT VARIANT 360 360 /FTID-VAR_007743.
 FT VARIANT 360 360 C -> G (IN DDS).
 FT VARIANT 366 366 /FTID-VAR_007744.
 FT VARIANT 366 366 R -> C (IN WT).
 FT VARIANT 366 366 /FTID-VAR_007745.
 FT VARIANT 366 366 R -> H (IN DDS).
 FT VARIANT 366 366 /FTID-VAR_007746.
 FT VARIANT 373 373 H -> Q (IN DDS).
 FT
 Query Match 0.7%; Score 9; DB 1; Length 449;
 Best Local Similarity 100.0%; Pred. No. 2.4;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 218 PAPPPEPP 226
 Db 58 PAPPPEPP 66
 DB
 RESULT 9
 ID WT1_MOUSE STANDARD; PRT; 449 AA.
 AC P22561;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE WILMS' TUMOR PROTEIN HOMOLOG.
 GN WT1 OR WT-1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91141522; PubMed=1671709;
 RA Buckler A.J., Pelletier J., Haber D.A., Glaser T., Housman D.E.,
 RT "Isolation, characterization, and expression of the murine Wilms'
 RT tumor gene (WT1) during kidney development.";
 RL Mol. Cell. Biol. 11:1707-1712(1991).
 CC -1- FUNCTION: POTENTIAL ROLE IN TRANSCRIPTIONAL REGULATION. RECOGNIZES
 CC AND BINDS TO THE DNA SEQUENCE 5'-CGCCCCCGC-3'.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- ALTERNATIVE PRODUCTS: TWO ALTERNATIVE SPLICE SITES EXISTS.
 CC -1- TISSUE SPECIFICITY: KIDNEY.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING KIDNEY DEVELOPMENT.
 CC -1- SIMILARITY: BELONGS TO THE EGR FAMILY OF C2H2-TYPE ZINC-FINGER
 CC PROTEINS.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; M55512; AAA40573.1; -
DR PIR; A39692; A39692.
DR TRANSPAC; T02351; -
DR MGI; 98968; Wt1.
DR InterPro; IPR000822; -
DR InterPro; IPR000976; -
DR Pfam; PF00096; zf-C2H2; 4.
DR PRINTS; PR00048; ZINC-FINGER.
DR PRINTS; PR00049; WILMSTUMOUR.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 4.
KW Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;
FT Transcription regulation; Alternative splicing; Anti-oncogene.
FT DOMAIN 28 83 PRO-RICH.
FT ZN_FING 323 438 ZINC_FINGERS.
FT ZN_FING 323 347 C2H2-TYPE.
FT ZN_FING 353 377 C2H2-TYPE.
FT ZN_FING 383 405 C2H2-TYPE.
FT ZN_FING 414 438 C2H2-TYPE.
FT VARSPLIC 250 266 MISSING (IN ISOFORM 2 AND ISOFORM 3).
FT VARSPLIC 408 410 MISSING (IN ISOFORM 2 AND ISOFORM 4).
SQ SEQUENCE 449 AA; 49246 MW; 962381E9C8D7A380 CRC64;

Query Match 0.7%; Score 9; DB 1; Length 449;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 PAPPPEPP 226
| | | | | | | | | |
DB 59 PAPPPEPP 67

RESULT 10
WT1_PIG STANDARD; PRT; 449 AA.
ID WT1_PIG
AC 062651;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE WILMS' TUMOR PROTEIN HOMOLOG.
GN WT1.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
OC NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LMD; TISSUE=Kidney;
RA MEDLINE=98267201; PubMed=9602131;
RA Tsurutani N., Oda H., Nakatsuru Y., Imai Y., Zhang S., Ueno Y.,
RA Ishikawa T.;
RT cDNA cloning and developmental expression of the porcine homologue
RT of WT1.";
RL Gene 211:215-220(1998).
CC -1- FUNCTION: POTENTIAL ROLE IN TRANSCRIPTIONAL REGULATION. RECOGNIZES
CC AND BINDS TO THE DNA SEQUENCE 5'-CGCCCCCGC-3'.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- ALTERNATIVE PRODUCTS: TWO ALTERNATIVE SPLICE SITES EXISTS.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING KIDNEY DEVELOPMENT.
CC -1- SIMILARITY: BELONGS TO THE EGR FAMILY OF C2H2-TYPE ZINC-FINGER
CC PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AB010969; BAA28147.1; -
DR InterPro; IPR000822; -
DR InterPro; IPR000976; -
DR Pfam; PF00096; zf-C2H2; 4.
DR PRINTS; PR00048; ZINC-FINGER.
DR PRINTS; PR00049; WILMSTUMOUR.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 4.
KW Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;
FT Transcription regulation; Alternative splicing; Anti-oncogene.
FT DOMAIN 28 83 PRO-RICH.
FT ZN_FING 323 438 ZINC_FINGERS.
FT ZN_FING 323 347 C2H2-TYPE.
FT ZN_FING 353 377 C2H2-TYPE.
FT ZN_FING 383 405 C2H2-TYPE.
FT ZN_FING 414 438 C2H2-TYPE.
FT VARSPLIC 249 265 MISSING (IN ISOFORM 2 AND ISOFORM 3) (BY
FT VARSPLIC 407 409 SIMILARITY).
FT VARSPLIC 407 409 MISSING (IN ISOFORM 2 AND ISOFORM 4) (BY
FT VARSPLIC 407 409 SIMILARITY).
SQ SEQUENCE 449 AA; 49166 MW; 9C3E57B96F5A7B3 CRC64;

Query Match 0.7%; Score 9; DB 1; Length 449;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 PAPPPEPP 226
| | | | | | | | | |
DB 58 PAPPPEPP 66

RESULT 11
SOX8_CHICK STANDARD; PRT; 470 AA.
ID SOX8_CHICK
AC P57074;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE TRANSCRIPTION FACTOR SOX-8.
GN SOX8.
OS Gallus gallus (chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20302574; PubMed=10842083;
RA Bell K.M., Western P.S., Sinclair A.H.;
RT "SOX8 expression during chick embryogenesis.";
RL Mech. Dev. 94:257-260(2000).
CC -1- FUNCTION: MAY PLAY A ROLE IN CENTRAL NERVOUS SYSTEM, LIMB AND
CC FACIAL DEVELOPMENT. MAY BE INVOLVED IN MALE SEX DETERMINATION.
CC BINDS THE CONSENSUS MOTIF 5'-[AT][AT]CA[AT]G-3'.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED IN THE EMBRYO.
CC -1- SIMILARITY: CONTAINS 1 HMG BOX.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AF228664; AAF73917.1; -
 KM Transcription regulation; DNA-binding; Nuclear protein.
 FT DNA_BIND 109 177 HMG BOX.
 SQ SEQUENCE 470 AA; 50830 MW; 020ADD008AFA4334 CRC64;

Query Match 0.7%; Score 9; DB 1; Length 470;
 Best Local Similarity 100.0%; Pred. No. 2.5;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 PGAPLGA 14
 CC |||||

DB 54 PGAPLGA 62

RESULT 12
 MAZ_HUMAN STANDARD; PRT; 477 AA.
 AC P56270; Q99443; Q15703;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE MYC-ASSOCIATED ZINC FINGER PROTEIN (MAZ1) (PURINE-BINDING
 DE TRANSCRIPTION FACTOR) (PUR-1) (ZF87) (ZF87).
 GN MAZ.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 OX [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92366479; PubMed=1502157;
 RA Bossone S.A., Asselin C., Patel A.J., Marcu K.B.;
 RT "MAZ, a zinc finger protein, binds to c-Myc and C2 gene sequences
 RT regulating transcriptional initiation and termination.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:7452-7456(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Carcinoma;
 RX MEDLINE=9232709; PubMed=1567856;
 RA Pyrc J.J., Moberg K.H., Hall D.J.;
 RT "Isolation of a novel cDNA encoding a zinc-finger protein that binds
 RT to two sites within the c-myc promoter.";
 RL Biochemistry 31:4102-4110(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pancratic islets;
 RX MEDLINE=96428591; PubMed=8831693;
 RA Teutsui H., Sakatsune O., Itakura K., Yokoyama K.K.;
 RT "Members of the MAZ family: a novel cDNA clone for MAZ from human
 RT pancreatic islet cells.";
 RL Biochem. Biophys. Res. Commun. 226:801-809(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96224025; PubMed=8626793;
 RA Parks C.L., Shenk T.;
 RT "The serotonin 1a receptor gene contains a TATA-less promoter that
 RT responds to MAZ and Sp1.";
 RL J. Biol. Chem. 271:4417-4430(1996).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lymphoblastoma;
 RX MEDLINE=9685418; PubMed=9685418;
 RA Song J., Murakami H., Tsutsui H., Tang X., Matsumura M., Itakura K.,
 RT Kanazawa I., Sun K., Yokoyama K.K.;
 RT "Genomic organization and expression of a human gene for Myc-
 RT associated zinc finger protein (MAZ).";
 RL J. Biol. Chem. 273:20603-20614(1998).
 CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR WITH DUAL ROLES
 CC IN TRANSCRIPTION INITIATION AND TERMINATION. BINDS TO TWO SITES,
 CC METAL AND METAL2, WITHIN THE C-MYC PROMOTER HAVING GREATER
 CC AFFINITY FOR THE FORMER. ALSO BINDS TO MULTIPLE G/C-RICH SITES
 CC WITHIN THE PROMOTER OF THE SPI FAMILY OF TRANSCRIPTION FACTORS.

CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
 CC -1- TISSUE SPECIFICITY: HEART, BRAIN, PLACENTA, LUNG, LIVER, SKELETAL
 CC MUSCLE, AND PANCREAS. SEEMS NOT TO BE EXPRESSED IN KIDNEY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M94046; -; NOT ANNOTATED CDS.
 CC EMBL; M93339; -; NOT ANNOTATED CDS.
 CC EMBL; D85131; BA12728.1; ALT_INIT.
 CC EMBL; U33819; AB04121.1; ALT_INIT.
 CC EMBL; AB017335; BAA33064.1; -.
 CC MIM; 600999; -.
 CC InterPro; IPR000822; -.
 CC Pfam; PF00096; zf-C2H2; 6.
 CC PRINTS; PR00048; ZINC_FINGER.
 CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
 CC PROSITE; PS0157; ZINC_FINGER_C2H2_2; 5.
 CC TRANSCRIPTION regulation; Zinc-finger; Metal-binding; DNA-binding;
 CC RNA-binding; Repeat; Nuclear protein.
 CC DOMAIN 190 413 ZINC_FINGERS.
 CC FT ZN_FING 190 212 C2H2-TYPE.
 CC FT ZN_FING 279 301 C2H2-TYPE.
 CC FT ZN_FING 307 329 C2H2-TYPE.
 CC FT ZN_FING 337 360 C2H2-TYPE.
 CC FT ZN_FING 366 388 C2H2-TYPE.
 CC FT ZN_FING 392 413 C2H2-TYPE.
 CC FT DOMAIN 96 108 POLY-ALA.
 CC FT DOMAIN 133 139 POLY-PRO.
 CC FT DOMAIN 157 161 POLY-ALA.
 CC FT DOMAIN 245 249 POLY-GLY.
 CC FT DOMAIN 435 449 POLY-ALA.
 CC FT CONFLICT 259 259 MISSING (IN REF. 3).
 CC FT CONFLICT 401 401 L -> M (IN REF. 2 AND 4).
 CC FT CONFLICT 443 447 MISSING (IN REF. 3).
 CC FT CONFLICT 443 447 MISSING (IN REF. 3).
 CC SQ SEQUENCE 477 AA; 48607 MW; C04C80F32C3C6825 CRC64;

Query Match 0.7%; Score 9; DB 1; Length 477;
 Best Local Similarity 100.0%; Pred. No. 2.5;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 218 PAPPPEPP 226
 CC |||||

DB 131 PAPPPEPP 139

RESULT 13
 PERL_VOICA STANDARD; PRT; 512 AA.
 ID PERL_VOICA
 AC P81131;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE PERPHORIN I PRECURSOR.
 OS Volvox carteri.
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 OC Volvocaceae; Volvox.
 OX NCBI_TaxID=3067;
 OX [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=93209229; PubMed=8458341;
 RA Sumper M., Berg E., Wenzl S., Godl K.;
 RT "How a sex pheromone might act at a concentration below 10(-16) M.";
 RL EMBL J. 12:831-836(1993).
 CC -1- FUNCTION: MAY BE INVOLVED IN CONVERSION OF ASEXUAL MALES AND
 CC FEMALES TO THE SEXUAL PATHWAY.
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.

```

CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X69801; CAB56808.1; -
KW Signal: Extracellular matrix; Glycoprotein; Multigene family.
FT SIGNAL 1
FT CHAIN 18
FT DOMAIN 192 203
FT CARBOHYD 49 49
FT CARBOHYD 96 96
FT CARBOHYD 118 118
FT CARBOHYD 378 378
FT CARBOHYD 381 381
FT CARBOHYD 403 403
FT CARBOHYD 476 476
SQ SEQUENCE 512 AA; 53878 MW; 091841C5D80F5E3 CRC64;

Query Match
Best Local Similarity 0.7%; Score 9; DB 1; Length 512;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 218 PAPPPEPP 226
Db 190 PAPPPEPP 198

RESULT 14
WASP_MOUSE
ID WASP_MOUSE STANDARD; PRT; 520 AA.
AC P70315;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 01-OCT-2000 (Rel. 40, Last annotation update)
DE WISKOTT-ALDRICH SYNDROME PROTEIN HOMOLOG (WASP).
GN WAS OR WASP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C;
RX MEDLINE=96115600; PubMed=8666397;
RA Derry J.M.J., Wiedemann P., Blair P., Wang Y., Kerns J.A.,
RA Lemaitre V., Godfrey V.L., Wilkinson J.E., Francke U.;
RT "The mouse homolog of the Wiskott-Aldrich syndrome protein (WASP)
RT gene is highly conserved and maps near the scurfy (sf) mutation on
RT the X chromosome.";
RL Genomics 29:471-477(1995).
CC -1- FUNCTION: POSSIBLE REGULATOR OF LYMPHOCYTE AND PLATELET FUNCTION.
CC MAY BE INVOLVED IN SIGNALING PATHWAYS WITH CYTOSKELETAL FUNCTION.
CC (BY SIMILARITY).
CC -1- DOMAIN: THE WH1 (WASP HOMOLOGY 1) DOMAIN MAY BIND A PRO-RICH
CC LIGAND.
CC -1- SIMILARITY: CONTAINS 1 WH1 DOMAIN.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U54788; AAC52556.1; -

```

```

DR MGD: MGI:105059; Wasp.
DR InterPro: IPR000095;
DR InterPro: IPR001960;
DR Pfam: PF00786; PBD; 1.
DR Pfam: PF00568; WH1; 1.
DR PROSITE: PS50108; GBD; 1.
KW Repeat.
FT DOMAIN 41 147
FT DOMAIN 240 259
FT REPEAT 354 363
FT REPEAT 393 402
FT DOMAIN 162 167
FT DOMAIN 314 321
FT DOMAIN 324 341
FT DOMAIN 368 373
FT DOMAIN 376 379
FT DOMAIN 384 390
FT DOMAIN 397 403
FT DOMAIN 408 424
FT DOMAIN 503 520
SQ SEQUENCE 520 AA; 54191 MW; 9C223733C59F0C8A CRC64;

Query Match
Best Local Similarity 0.7%; Score 9; DB 1; Length 520;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 218 PAPPPEPP 226
Db 409 PAPPPEPP 417

RESULT 15
VP61_NPVAC
ID VP61_NPVAC STANDARD; PRT; 543 AA.
AC Q03209;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE 61 KDA PROTEIN.
GN P61.
OS Autographa californica nuclear polyhedrosis virus (ACNPV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolydnavirus.
OC NCBI_TaxID=46015;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-E2;
RX MEDLINE=93138405; PubMed=1487149;
RA Pham D.O.-D., Sivasubramanian N.;
RT "Sequence and in vitro translational analysis of a 1629-nucleotide
RT ORF in Autographa californica nuclear polyhedrosis virus strain E2."
RL Gene 122:345-348(1992).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-C6;
RX MEDLINE=92024079; PubMed=1926775;
RA Possee R.D., Sun T.P., Howard S.C., Ayres M.D., Hill-Perkins M.,
RA Gearing R.L.;
RT "Nucleotide sequence of the Autographa californica nuclear
RT polyhedrosis 9.4 kbp EcoRI-I and -R (polyhedrin gene) region.";
RL Virology 185:229-241(1991).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN-C6;
RX MEDLINE=94303173; PubMed=8030224;
RA Ayres M.D., Howard S.C., Kuzio J., Lopez-Ferber M., Possee R.D.;
RT "The complete DNA sequence of Autographa californica nuclear
RT polyhedrosis virus.";
RL Virology 202:586-605(1994).
CC -1- FUNCTION: PROBABLY PLAYS AN IMPORTANT ROLE IN THE PERSISTENCE AND
CC SURVIVAL OF THE VIRUS. MAY BE A STRUCTURAL COMPONENT IN THE
CC ENVELOPE OF THE POLYHEDRON OR THE ENVELOPE OF THE POLYHEDRA-

```

CC DERIVED VIRUS (PDV).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z11662; CAA77730.1; -;
 DR EMBL: M75679; AAA46705.1; -;
 DR EMBL: L22858; AAA66639.1; -;
 DR PIR: S25128; S25128.
 DR PIR: H40781; H40781.
 DR DOMAIN 143 320 PRO-RICH.
 FT DOMAIN 278 295 POLY-PRO.
 FT CONFLICT 361 361 V -> A (IN REF. 1).
 SO SEQUENCE 543 AA; 60713 MW; 196D295D5E5FC100 CRC64;

Query Match 0.7%; Score 9; DB 1; Length 543;
 Best Local Similarity 100.0%; Pred. No. 2.8;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 218 PAPPPIPP 226
 |||||
 Db 195 PAPPPIPP 203

Search completed: July 19, 2001, 16:11:46
 Job time: 215 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2001, 16:04:20 ; Search time 27.58 Seconds

(without alignments)
2692,690 Million cell updates/sec

Title: us-09-501-171a-4

Perfect score: 1225

Sequence: 1 MFARKPPGAAPLCAMPYDQ.....PYSLENYTSHPSPDSMW 1225

Scoring table: OLIGO

Searched: 412676 seqs, 60623988 residues

Word size: 0

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: listing first 45 summaries

Database: A_Geneseq_0601.*

1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1225	100.0	1225	21	A human neural pla
2	876	71.5	1084	20	AAW23900
3	868	70.9	1040	18	AAW24559
4	381	31.1	756	19	AAW60664
5	284	23.2	1247	21	AAW07974
6	26	2.1	1192	20	AAW23899
7	26	2.1	1211	18	AAW24560
8	26	2.1	1211	21	AAW2336
9	11	0.9	30	21	AAW36465
10	9	0.7	11	21	AAW8346
11	9	0.7	11	22	AAW30628

12	9	0.7	15	18	AAW39024
13	9	0.7	18	17	AAW05474
14	9	0.7	21	22	AAW30656
15	9	0.7	23	22	AAW30657
16	9	0.7	36	21	AAW53868
17	9	0.7	130	22	AAW29756
18	9	0.7	210	16	AAW85066
19	9	0.7	210	18	AAW22883
20	9	0.7	429	16	AAW85065
21	9	0.7	449	19	AAW47176
22	9	0.7	449	21	AAW98804
23	9	0.7	449	21	AAW98805
24	9	0.7	449	21	AAW80196
25	9	0.7	449	21	AAW80197
26	9	0.7	456	20	AAW17067
27	9	0.7	519	19	AAW70285
28	9	0.7	520	22	AAW49336
29	9	0.7	539	19	AAW48358
30	9	0.7	600	20	AAW82660
31	9	0.7	601	21	AAW79240
32	9	0.7	606	21	AAW79241
33	9	0.7	647	11	AAW60001
34	9	0.7	797	19	AAW69373
35	9	0.7	851	19	AAW69370
36	9	0.7	859	19	AAW69372
37	9	0.7	859	19	AAW69374
38	9	0.7	871	19	AAW69375
39	9	0.7	897	19	AAW69371
40	9	0.7	928	11	AAW05305
41	9	0.7	928	11	AAW05289
42	9	0.7	928	14	AAW36534
43	9	0.7	928	16	AAW74271
44	9	0.7	928	16	AAW71680
45	9	0.7	928	16	AAW71681

ALIGNMENTS

RESULT 1	
ID	AAW07973 standard; Protein; 1225 AA.
XX	
AC	AAW07973;
XX	
DT	14-NOV-2000 (first entry)
XX	
DE	A human neural plakophilin related armidillo protein.
XX	
KW	Human; presenilin; PS1; PS2; nerve cell growth; nerve damage; hNPRAP;
KW	neural plakophilin related armidillo protein; Alzheimer's disease;
KW	Parkinson's disease; amyotrophic lateral sclerosis; ALS; stroke;
KW	multiple sclerosis; ischemia; stroke; neural paropathy; diabetic crush;
KW	motor neuron disease; peripheral neuropathy; neuropathy; diabetes;
KW	spinal cord injury; facial nerve crush.
XX	
OS	Homo sapiens.
XX	
PN	W0200047615-A2.
PD	
PD	17-AUG-2000.
XX	
PF	11-FEB-2000; 2000MO-CA00126.
XX	
PR	12-FEB-1999; 9905-0119835.
XX	
PA	(UTOR) UNIV TORONTO GOVERNING COUNCIL.
XX	
PI	St George-Hyslop PH, Fraser PE;
XX	
DR	WPI; 2000-524531/47.
XX	
DR	N-PSDB; AAAS9700.
XX	

Peptide resembling
SH3-binding peptid
Trypsin modulating
Trypsin modulating
Human colon cancer
Rice glutaredoxin
Wilms' tumour anti
Wilms' tumour prot
Wilms' tumour WT1
Human WT1 protein
Mouse WT1 protein
Mouse Wilms' tumou
Human Wilms' tumou
Human 3-OST-4 prot
Oryza sativa RNA p
Murine WASP protei
Fractionated X irr
Caulliflower L-gala
Human diarylsulfon
Human diarylsulfon
Insect receptor po
Modified retinobla
Modified retinobla
Modified retinobla
Modified retinobla
Modified retinobla
Cancer suppressin
Predicted retinobl
Retinoblastoma (R
Retinoblastoma tum
Recombinant ppl10R

PT stimulation of nerve cell growth using human Neural Plakophilin Related
 PT Armadillo Protein (hNPRAP) polypeptide. hNPRAP interacts with presenilins
 PT diseases such as Alzheimer's, Parkinson's, and stroke -
 XX
 XX Claim 3; Page 24-27; 33pp; English.
 CC The present sequence represents human Neural Plakophilin Related
 CC Armadillo Protein (hNPRAP) polypeptide. hNPRAP interacts with presenilins
 CC (PS1 and PS2). The specification describes a method
 CC for stimulating the growth of nerve cells, comprising contacting
 CC them with hNPRAP. The hNPRAP polypeptide and polynucleotide are useful
 CC for treating nerve damage caused by a variety of diseases or physical
 CC traumas, including Alzheimer's disease, Parkinson's disease, amyotrophic
 CC lateral sclerosis (ALS), multiple sclerosis, stroke, ischemia associated
 CC with stroke, neural paropathy, motor neuron diseases, sciatic crush,
 CC peripheral neuropathy, neuropathy associated with diabetes, spinal cord
 CC injuries and facial nerve crush.
 XX
 XX Sequence 1225 AA;
 SQ
 Query Match 100.0%; Score 1225; DB 21; Length 1225;
 Best Local Similarity 100.0%; Pred. NO. 0;
 Matches 1225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MFARKPGAAPLGAMPVDPQPSASEKTSLSPLGINTSNGDSEETTSALIASVKEOEL 60
 DB 1 mfarppgaaplgampvdpqpsasektslsplgintsgdsettsaliasvkeqel 60
 QY 61 OPERITRELEAROTIVASOLEERKIGSETGSMSSKSAEBOFOMOSODGKIDELITTG 120
 DB 61 qferitrelaeerqivaqslcerkigsetgsmssaseeqfqwsgdqgkidelitg 120
 QY 121 LELVSCIRISOESGILPDYSTERGERPSILSOSALONSKPEGSFOYPATYHSNOTLAL 180
 DB 121 lelvsclrsiqesgilpdystergerpsllsqsaldnsgbegstqypasynqclal 180
 QY 181 GETTPSQLPARQTQARATGOSFSOGTTSRAGHLGPEPAPPPPPPPPPPPPPPPPPPP 240
 DB 181 gettpsqipargtgaratgsgfsqgtstsraghlagpepppppppppppppppppp 240
 QY 241 PDAPRAAAAAALYSSSTSLPAPRGGSGPLAPOGSGSPKLRGSGAPBAGATYAAPRGSSP 300
 DB 241 pdapraaaaaalysstslpapprggsplaaogsgspkllrggsapagatyaaprgssp 300
 QY 301 KOSPERLAKSYSTSPINIVVSAGLSPIRTSPPTVOSTISSPPIHOLSTIGTYATLS 360
 DB 301 kgsperlaksystspiniivvsaglsplirvspptvostissppiholstigtatls 360
 QY 361 PTKRLVHASEOYSKHSELYATATLQRPGLAAGSRASYSQHGHLGDELALQSPEHHI 420
 DB 361 ptkrlvhasseqskhsqelyatatlqrgpslaagsraaysqghlqdelalqspehhi 420
 QY 421 DPIYEDRYOKRPMKSTISOSGDPRLPRAHNTGTYRTSTPSSPGVDSVLAQOTGSOHQPON 480
 DB 421 dpiyedryokrpmkstisosgdprrpahnrtgtyrtstpspgvdsvplqctlgshqpna 480
 QY 481 AAAATFORASTAAGPASNADPYROLQCPSEVESBYSKGPALPREGTLASPSIDSIQK 540
 DB 481 aaaatfgrastaaagpasnadyrqlqycpsevesbyskgsplapregtlarspsidsiqk 540
 QY 541 DPREFGWRDPELPEYIQMLQHOFPVSVOGNAAYLQHLCEGNNKTKAETIRGOGIQLVLDL 600
 DB 541 dprefgwrdepelpeyiqmlqhofpsvognaaaylqhlcefgnnkkaetirrgogiqllvdl 600
 QY 601 LDHMTFVHRSGAGLRLVYGGKANDDKIKLKNCGGIPALVRLRKTTDLEIELVTVGV 660
 DB 601 ldhmtfvrsgaglrlyvgkanddkialknogipalvrlrkttdeleivtvgv 660
 QY 661 LWNLSSCDALKMPITODALAVLTNAVILPHSGWENSPLODDRKTQLSHSSQVLRNATGCLR 720
 DB 661 lwnlsscdalkmpitodalavltnavilphsgwensploddrktqlshssqvlrnatgclr 720

QY 721 NVSSAGEARRRRRRECDGLTLDALLVYIOSALGSSSIDSKTYEENCVCILRNISYLAETS 780
 DB 721 nvssagearrrrrrecdgltldallvyiosalgsssidsktyeencvcilrnisyriaets 780
 QY 781 QCGHMTDDELDELGCSEANGKRAESSGCGKRRKKKSSODDWDGVPPLDCAEPKGIQM 840
 DB 781 qcghmtddeldelgcseangkraessgcgkrrkkkssoddwdgvppldcaepkgiqm 840
 QY 841 LWHPSIVKRYTLTLSCSNPDLEAGACALONLWAGSMKWSYVIRAAVREKKGPILEVL 900
 DB 841 lwhpsivkrytltlscsnpdleagacalonaagsmkwsyviraaarkkgpillel 900
 QY 901 LRINDRVYCAVATALRNALDVYRNKELIGYAMRDVYHRLPGNNSNNTASKMSDDTV 960
 DB 901 lrindrvycavatarnaldvyrnkelligyamrdvyhrlpgnnsnntaskmsddtv 960
 QY 961 TANCCILHEVITRKNNMNAKALRDAGIEKLVTGITSKSGDKHSPKYVKAASQVLSMMQYR 1020
 DB 961 tawccilhevitrknnmnakalrdagiekltvgitsksgdkhspkvkaasqvlsmmqyr 1020
 QY 1021 DLRSILKKDQMSQYHFAVASSSTIERDRQRPYSSSRTPSISPYRVSPNNRSASAPAPREM 1080
 DB 1021 dlrsilkkdqwsgyhfavassstierdrqrpysstpsispyrvspnnrsasapaprem 1080
 QY 1081 ISLKERKTDYECTGSNATYHGAKEHTSRKDMATQNTGISTLYNNSYGAPADIKHNOY 1140
 DB 1081 islkerktdyecnatygahgakehtsrkdmatqntgisltylnsygapedikhnoy 1140
 QY 1141 SAQPVQEPSRKDYETVQFONSTRYDESPEDOVHHRPPASEYTMHLGKSTGNYDF 1200
 DB 1141 saqpvqepsrkyetvqfonsrtydespedovhhrppaseytmhlgkstgnydf 1200
 QY 1201 YSARPYSEINETSHPASPSMW 1225
 DB 1201 ysaarpyseinyetshypaspsmw 1225

RESULT 2
 AAY23900
 ID AAY23900 standard; protein; 1084 AA.
 XX
 AC AAY23900;
 DT 27-SEP-1999 (first entry)
 DE Human resenilin binding armadillo protein GT24/hNPRAP.
 XX
 DE Human: presenilin 1; PS1; presenilin-binding protein; interacting domain;
 KW presenilin allele; Alzheimer's disease; senile dementia;
 KW psychiatric disease; schizophrenia; depression; neurological disease;
 KW stroke; cerebral haemorrhage; p0071; armadillo protein.
 OS Homo sapiens.
 XX
 FH Key location/Qualifiers
 FT Misc-difference 115
 FT /note= "not specified"
 FT
 MO935501-A1.
 PN 15-JUL-1999.
 PD 08-JAN-1999; 99WO-CA00018.
 PF 09-JAN-1998; 98US-0070948.
 PR (UTOR) UNIV TORONTO GOVERNING COUNCIL.
 PA Fraser PE, St George-Hyslop PH;
 PI WPI; 1999-419410/35.
 XX Identifying substances that alter presenilin interactions, useful
 PT

XX Claim 1; Page 99-101; 133pp; English.

ES
XX
CC GT24, a human protein with 'armadillo' repeats, has been identified
CC as a presenilin-interacting protein (PIP). A yeast two-hybrid kit
CC was employed to screen a human brain cDNA library for clones which
CC interacted with presenilin transmembrane 6-7 loop domain; mutations
CC in this loop domain are known to be causative of Alzheimer's
CC disease (AD). 9 PIP gene sequences (see AAT79966-74) including GT24
CC were identified. PIP nucleic acids, PIP proteins and peptides
CC (especially the presenilin interacting domain), antibodies to PIPs,
CC altered with PIP nucleic acids can be used for the diagnosis,
CC therapy and study of AD and related disorders. They can be used to
CC identify compounds which can modulate the expression of a PIP gene
CC or which bind to a PIP or modulate its activity.

XX
SQ Sequence 1040 AA;

Query Match 70.9%; Score 868; DB 16; Length 1040;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 968; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 257 STLPAPRRGSPPLAPGSGPTKLRGSAPEGATYAAPRSGSPKSPSLAKSYSTSSP 316
DB 72 STLPAPRRGSPPLAPGSGPTKLRGSAPEGATYAAPRSGSPKSPSLAKSYSTSSP 131
QY 317 INIVSSAGLSFIRVTSPTVOSTISSPILHQLSTICTVATLSPTKRLVHASRQYSKHS 376
DB 132 INIVSSAGLSFIRVTSPTVOSTISSPILHQLSTICTVATLSPTKRLVHASRQYSKHS 191
QY 377 OELVATATLORPGSLAASRASYSOSGHILGPELRALOSPEHHTIDPIYEDVYVOKPQRS 436
DB 192 GELVATATLORPGSLAASRASYSOSGHILGPELRALOSPEHHTIDPIYEDVYVOKPQRS 251
QY 437 LSQSGDPLPRAHTGTTRTSTAPSSPGVSVPLQRTGSGHQPONAAATFORASYAAGPA 496
DB 252 LSQSGDPLPRAHTGTTRTSTAPSSPGVSVPLQRTGSGHQPONAAATFORASYAAGPA 311
QY 497 SNYAPRYQLQYCPVESPYSKSGPALPPEGILARSPSIDSIQKDPREFGRMDPELPVT 556
DB 312 SNYAPRYQLQYCPVESPYSKSGPALPPEGILARSPSIDSIQKDPREFGRMDPELPVT 371
QY 557 OMLQHOFPVSQSNAAAYLOHLCFGDNKIKAEIRROGIGIOLVLDLDRHMTVHHSACAL 616
DB 372 OMLQHOFPVSQSNAAAYLOHLCFGDNKIKAEIRROGIGIOLVLDLDRHMTVHHSACAL 431
QY 617 RNLVYGGKANDDKIALKNCGGIPALVRLRKTDTLEIRELVTVGLMNLSSCDALKMPIIQ 676
DB 432 RNLVYGGKANDDKIALKNCGGIPALVRLRKTDTLEIRELVTVGLMNLSSCDALKMPIIQ 491
QY 677 DALAVITNATVITPHSGWENSPLODDRKIOHSSQVLRNATGCLRNVSAGSEARRMREC 736
DB 492 DALAVITNATVITPHSGWENSPLODDRKIOHSSQVLRNATGCLRNVSAGSEARRMREC 551
QY 737 DELTALLVYIOSAGSSEISDKTENCVCILRNLSYRLAETSOGQHMGTDELDELGLCG 796
DB 552 DELTALLVYIOSAGSSEISDKTENCVCILRNLSYRLAETSOGQHMGTDELDELGLCG 611
QY 797 EANGDAESSGCGWKKKKKKSSQDQWGVPLPDCAEPPKGIQIMHPISYKPYITLISE 856
DB 612 EANGDAESSGCGWKKKKKKSSQDQWGVPLPDCAEPPKGIQIMHPISYKPYITLISE 671
QY 857 CSNPPTLEGAGALONLAAGSWMKSVYIRAAVRKEGIPITVELIRITNDVQCAVATAL 916
DB 672 CSNPPTLEGAGALONLAAGSWMKSVYIRAAVRKEGIPITVELIRITNDVQCAVATAL 731
QY 917 RNMAIDVNRKELIGRYAMADLVHRLPGGNNNSNTASKMSMDTAVACCTLHEVITKME 976
DB 732 RNMAIDVNRKELIGRYAMADLVHRLPGGNNNSNTASKMSMDTAVACCTLHEVITKME 791
QY 977 NAKALRDAGGIEKLVGISKSGDKHSPPVYKAAISOVLNSMWOYRDLRLSLYKKDKGMSQYHF 1036

DB 792 NAKALRDAGGIEKLVGISKSGDKHSPPVYKAAISOVLNSMWOYRDLRLSLYKKDKGMSQYHF 851
QY 1037 VASSSTIERDRORPYSSTRPSISPVVSPNNRSASAPSPREMI SLKERKTDYECTGSN 1096
DB 852 VASSSTIERDRORPYSSTRPSISPVVSPNNRSASAPSPREMI SLKERKTDYECTGSN 911
QY 1097 ATYHGAKGEITSKDKMTAQNIGISITLNRNSYGAPREDIKHNOVSNOQPPQPSKDYET 1156
DB 912 ATYHGAKGEITSKDKMTAQNIGISITLNRNSYGAPREDIKHNOVSNOQPPQPSKDYET 971
QY 1157 YQFONSTRYDPSFEDQVHHRPPASEXTMILGLKSTGNVYDFSAARPYSELNVTSH 1216
DB 972 YQFONSTRYDPSFEDQVHHRPPASEXTMILGLKSTGNVYDFSAARPYSELNVTSH 1031
QY 1217 YPASPSWV 1225
DB 1032 YPASPSWV 1040

RESULT 4

AAW60664
ID AAW60664 standard; Protein; 756 AA.

XX
AC AAW60664;
XX
DT 26-OCT-1998 (first entry)

XX
DE Human ALARM protein.

XX
KW ALARM; adherens-junction linked arm protein; delta-catenin;
KM Presenilin 1; diagnosis; Alzheimer's disease; antibody; human.

XX
OS Homo sapiens.

XX
FH Key

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

WO9825142-A1.
11-JUN-1998.
02-DEC-1997; 97WO-US22093.
02-DEC-1996; 96US-0031556.
(BGM) BRIGHAM & WOMENS HOSPITAL.
Kosik KS, Zhou J;
WPI; 1998-333494/29.
N-PSDB; AAV38305.
Human ALARM polypeptide(s) binding to presenilin 1 - useful, e.g. to detect presenilin 1 and diagnose human diseases caused by mutant form of protein interacting with ALARM
Claim 4; Fig 1A-F; 62pp; English.
This novel human protein, named ALARM or delta-catenin, is able to bind to presenilin 1. Its amino acid sequence was deduced from a cDNA clone (see AAV38305) isolated from a human brain cDNA library. ALARM (named for adherens-junction linked arm protein) is expressed almost exclusively in brain tissue. It contains 4 copies of the arm repeat that is characteristic of catenin proteins. Overall, ALARM and delta-catenin are 52.3% similar and 32.1% identical. The

CC invention provides isolated ALARM polypeptide, nucleic acids,
CC vectors, host cells and specific antibodies. Presenilin 1 protein
CC can be detected in samples (especially human cerebrospinal fluid),
CC by contacting with ALARM (claimed). Mutation of the presenilin 1
CC gene is associated with familial Alzheimer's disease. ALARM can be
CC used to diagnose human diseases caused by a mutant form of a protein
CC interacting with ALARM, by analysing fluid samples to detect the
CC ALARM-interacting protein (claimed). ALARM can also be used to
CC diagnose altered levels of presenilin 1, by contacting a sample with
CC ALARM and determining bonding (claimed). ALARM polypeptides and
CC antibodies can be used to detect an ALARM-containing complex in
CC biological samples, by contacting with polypeptide or antibody and
CC determining whether the polypeptide/antibody binds to a sample
CC component (claimed). The antibodies are also useful to detect ALARM
CC polypeptides and to measure the effect of candidate compounds on
CC expression or localisation of ALARM. They are useful to evaluate
CC engineered cells prior to introduction in gene therapy, to inhibit
CC abnormal ALARM activity or to generate anti-idiotypic antibodies,
CC useful therapeutically.

XX Sequence 756 AA:

Query Match 31.1%; Score 381; DB 19; Length 756;

Best Local Similarity 99.6%; Pred. No. 0;

Matches 681; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 444 PLPAPHTGTRTSTAPSSPEVDSPVLQRTGSGHGPONMAAATFORASVAGPASNADPY 503
DB 11 PLPAPHTGTRTSTAPSSPEVDSPVLQRTGSGHGPONMAAATFORASVAGPASNADPY 70
QY 504 RQIQCPVSPVSPKSGPALPPEGTARSPSIDSIKDPREGWRDPELVYQMLQHF 563
DB 71 RQIQCPVSPVSPKSGPALPPEGTARSPSIDSIKDPREGWRDPELVYQMLQHF 130
QY 564 PSYOSNAALYQHLCEGDNKIRKAEIRROGIGOLLVDLHBMTEVHRSAGALRNLYVK 623
DB 131 PSYOSNAALYQHLCEGDNKIRKAEIRROGIGOLLVDLHBMTEVHRSAGALRNLYVK 190
QY 624 ANDNRIKALNCGIPALVRLRLKRTDLEIRELVTVGNLNSCDALMKPIIODALAVLT 683
DB 191 ANDNRIKALNCGIPALVRLRLKRTDLEIRELVTVGNLNSCDALMKPIIODALAVLT 250
QY 684 NATIRHSGWENPIODDKRIQIHSQVLRNATGCLRNSAGEEKRMRREDCGTLTDL 743
DB 251 NATIRHSGWENPIODDKRIQIHSQVLRNATGCLRNSAGEEKRMRREDCGTLTDL 310
QY 744 LVIYOSALGSEIDSKTVENCVCILRNLSYRLAETSGOHHMGTDELDCGLCEANGKDA 803
DB 311 LVIYOSALGSEIDSKTVENCVCILRNLSYRLAETSGOHHMGTDELDCGLCEANGKDA 370
QY 804 ESSGCGWKKKKKKSSODQMDVGPRLDCAEPKGIOMLWHPSTIVKPYTLTLESCSNPDTL 863
DB 371 ESSGCGWKKKKKKSSODQMDVGPRLDCAEPKGIOMLWHPSTIVKPYTLTLESCSNPDTL 430
QY 864 EGAAGALQNLAAAGSMKSYVYIRAAVKEKGLPILVELLRLINDRVYCAVATLRNALDY 923
DB 431 EGAAGALQNLAAAGSMKSYVYIRAAVKEKGLPILVELLRLINDRVYCAVATLRNALDY 490
QY 924 RNKLGLGVAMRLVHRLFGNNSNNTASKAMSDDTVTAVCCTLHVIYIRKNNENAKALRD 983
DB 491 RNKLGLGVAMRLVHRLFGNNSNNTASKAMSDDTVTAVCCTLHVIYIRKNNENAKALRD 550
QY 984 AGGIEKLGVISKSGDKHSKPVVKAASQVLSNKMWOYRDLRLSLYKKDKGWSQYHFAVSSSTI 1043
DB 551 AGGIEKLGVISKSGDKHSKPVVKAASQVLSNKMWOYRDLRLSLYKKDKGWSQYHFAVSSSTI 610
QY 1044 ERROBPYSSTSPSTSPYRVSPNNSASAPASPRMISLKERKTYECTGSAATYHGA 1103
DB 611 ERROBPYSSTSPSTSPYRVSPNNSASAPASPRMISLKERKTYECTGSAATYHGA 670
QY 1104 GEHTRKAMTAONTGISTLYRNS 1127
DB 1104 GEHTRKAMTAONTGISTLYRNS 1127

DB 671 gehtrkamdtaontgistlyrns 694

RESULT 5

AB07974 AAB07974 standard; Protein: 1247 AA.

AC AAB07974;

DE 14-NOV-2000 (first entry)

DE A murine neural plakophilin related armidillo protein.

XX Human; presenilin; PS1; PS2; nerve cell growth; nerve damage; hNRPAP;
KW neural plakophilin related armidillo protein; Alzheimer's disease;
KW Parkinson's disease; amyotrophic lateral sclerosis; ALS; stroke;
KW multiple sclerosis; ischemia; stroke; neural paropathy; sciatic crush;
KW motor neuron disease; peripheral neuropathy; neuropathy; diabetes;
KW spinal cord injury; facial nerve crush.

OS Mus sp.

PN W0200047615-A2.

PD 17-AUG-2000.

PF 11-FEB-2000; 2000WO-CA00126.

PR 12-FEB-1999; 99US-0119835.

PA (UTOR) UNIV TORONTO GOVERNING COUNCIL.

PI St George-Hyslop PH, Fraser PE;

DR WPI: 2000-524531/47.

XX N-PSDB; AAS59701.

PT Stimulation of nerve cell growth using human Neural Plakophilin Related

PT Armidillo Protein (hNRPAP) polypeptide, useful for the treatment of

PT diseases such as Alzheimer's, Parkinson's, and stroke

PS Disclosure: Page 30-33; 33pp; English.

XX The present sequence represents Neural Plakophilin Related Armidillo
CC Protein (NRPAP) polypeptide. Human NRPAP interacts with presenilins
CC (PS1 and PS2). The specification describes a method
CC for stimulating the growth of nerve cells, comprising contacting
CC them with hNRPAP. The hNRPAP polypeptide and polynucleotide are useful
CC for treating nerve damage caused by a variety of diseases or physical
CC traumas, including Alzheimer's disease, Parkinson's disease, amyotrophic
CC lateral sclerosis (ALS), multiple sclerosis, stroke, ischemia associated
CC with stroke, neural paropathy, motor neuron diseases, sciatic crush,
CC peripheral neuropathy, neuropathy associated with diabetes, spinal cord
CC injuries and facial nerve crush.

SO Sequence 1247 AA:

Query Match 23.2%; Score 284; DB 21; Length 1247;

Best Local Similarity 100.0%; Pred. No. 3.6e-243;

Matches 284; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 514 SPYKSGPALPPEGTARSPSIDSIKDPREGWRDPELVYQMLQHPPEVOSNAAY 573
DB 511 SPYKSGPALPPEGTARSPSIDSIKDPREGWRDPELVYQMLQHPPEVOSNAAY 570
QY 574 LOHLCFGDNKIRKAEIRROGIGOLLVDLHBMTEVHRSAGALRNLYYKANDNKIALK 633
DB 571 LOHLCFGDNKIRKAEIRROGIGOLLVDLHBMTEVHRSAGALRNLYYKANDNKIALK 630
QY 634 NCGGIPALVRLRLKRTDLEIRELVTVGNLNSCDALMKPIIODALAVLNNAVLIIPSGW 693
DB 631 NCGGIPALVRLRLKRTDLEIRELVTVGNLNSCDALMKPIIODALAVLNNAVLIIPSGW 690

OY 694 ENSPLDDEK10LHSSOVLNRNATGCLRNVSAGEARRRMRPCDGLTDALLVYIOSALGS 753
 DB 691 ensplqddk1qihssqylnatgclrnvsagearrmrccgldallvylqsalgs 750
 OY 754 SEIDSKTVENCVCILRNLSYRLAETSGOQHMGTDLDGLGCE 797
 DB 751 seidsktvencvcllnlsyrlaetsgqgmgtldelgdlcge 794

RESULT 6
 ID AAY23899 standard; protein; 1192 AA.
 AC AAY23899;
 DT 27-SEP-1999 (first entry)
 DE Human resenilin binding armadillo protein p0071.
 XX
 KM Human; presenilin 1; PS1; presenilin-binding protein; interacting domain;
 KM presenilin allele; Alzheimer's disease; senile dementia;
 KM psychiatric disease; schizophrenia; depression; neurological disease;
 KM stroke; cerebral haemorrhage; p0071; armadillo protein.
 OS Homo sapiens.
 XX
 PN MO9935501-AL.
 XX
 PD 15-JUL-1999.
 XX
 PF 08-JAN-1999; 99WO-CA00018.
 XX
 PR 09-JAN-1998; 98US-0070948.
 XX
 PA (UTOR) UNIV TORONTO GOVERNING COUNCIL.
 PI Fraser PE, St George-Hyslop PH;
 XX
 DR WPI; 1999-419410/35.
 XX
 PT Identifying substances that alter presenilin interactions, useful
 PT for screening individuals for presenilin alleles associated with
 PT Alzheimer's disease - useful for diagnosis of Alzheimer's disease
 PS
 PS Claim 10; Fig 1; 40pp; English.
 CC The present sequence represents a human presenilin-binding protein,
 CC termed p0071. It is a member of the armadillo family of proteins.
 CC The specification describes a method for identifying substances
 CC that alter the interaction of a presenilin with a presenilin-binding
 CC protein. The method comprises contacting the interacting domain of a
 CC presenilin protein to a presenilin-binding protein in the presence of
 CC a test substance, and measuring the interaction of the presenilin and
 CC the presenilin-binding protein. The method can be used to screen
 CC individuals for presenilin alleles associated with Alzheimer's disease
 CC and related disorders, such as senile dementia's, psychiatric diseases
 CC such as schizophrenia and depression, and neurological disease, such as
 CC stroke and cerebral haemorrhage.
 XX
 SQ Sequence 1192 AA;

Query Match 2.1%; Score 26; DB 20; Length 1192;
 Best Local Similarity 100.0%; Pred. No. 3.3e-14;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 913 ATALRNALDVNRKELIGRYAMRDIV 938
 DB 893 atalrmaidvnrnkeliqkyamrdiv 918

RESULT 7

AAW24560
 ID AAW24560 standard; protein; 1211 AA.
 AC AAW24560;
 XX
 DT 06-FEB-1998 (first entry)
 DE Presenilin-interacting protein p0071.
 XX
 KM Presenilin-interacting protein; human; Alzheimer's disease;
 KM diagnosis; therapy; transgenic animal; animal model; p0071.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Domain 509..1022
 FT /note="presenilin-interacting domain"
 XX
 PN MO9727296-AL.
 XX
 PD 31-JUL-1997.
 XX
 PF 27-JAN-1997; 97WO-CA00051.
 XX
 PR 02-JAN-1997; 97US-0034590.
 PR 26-JUL-1996; 96US-0592541.
 PR 05-JUL-1996; 96US-0021673.
 PR 12-JUL-1996; 96US-0021700.
 PR 08-NOV-1996; 96US-0029895.
 XX
 PA (HSCR-) HSC RES & DEV LP.
 PA (UTOR) UNIV TORONTO GOVERNING COUNCIL.
 PI Fraser PE, Rommens JW, St George-Hyslop PH;
 XX
 DR WPI; 1997-393684/36.
 DR N-PSDB; AAT79968.
 XX
 PT Presenilin-interacting protein genes - used to develop products for
 PT the diagnosis, therapy and study of Alzheimer's disease and related
 PT disorders
 PS
 PS Claim 1; Page 104-107; 133pp; English.
 CC p0071, a human protein with 'armadillo' repeats, has been identified
 CC as a presenilin-interacting protein (PIP). A yeast two-hybrid kit
 CC was employed to screen a human brain cDNA library for clones which
 CC interacted with presenilin transmembrane 6-7 loop domain; mutations
 CC in this loop domain are known to be causative of Alzheimer's
 CC disease (AD). 9 PIP gene sequences (see AAT79966-74) including p0071
 CC were identified. 9 PIP gene sequences (PIP proteins and peptides
 CC (especially the presenilin interacting domain), antibodies to PIPs,
 CC cells transformed with PIP nucleic acids, and transgenic animals
 CC altered with PIP nucleic acids can be used for the diagnosis,
 CC therapy and study of AD and related disorders. They can be used to
 CC identify compounds which can modulate the expression of a PIP gene
 CC or which bind to a PIP or modulate its activity.
 XX
 SQ Sequence 1211 AA;

Query Match 2.1%; Score 26; DB 18; Length 1211;
 Best Local Similarity 100.0%; Pred. No. 3.3e-14;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 913 ATALRNALDVNRKELIGRYAMRDIV 938
 DB 893 atalrmaidvnrnkeliqkyamrdiv 918

RESULT 8
 ID AAY92336 standard; protein; 1211 AA.

DT 16-FEB-2001 (first entry)
 XX Trypsin Modulating Coostatic Factor.
 DE
 XX Pesticidal: mosquito; Trypsin Modulating Coostatic Factor;
 KW TMOF; hormone; trypsin biosynthesis; digestive enzyme.
 XX
 OS Aedes aegypti.
 XX
 PN WO200063235-A2.
 XX
 PD 26-OCT-2000.
 XX
 PF 18-APR-2000; 2000WO-US10375.
 XX
 PR 21-APR-1999; 99US-0296113.
 XX
 PA (UYFL) UNIV FLORIDA RES FOUND INC.
 XX
 PI Borovsky D;
 XX
 DR WPI; 2000-687157/67.
 XX
 XX Recombinant plant cell transformed to express a polynucleotide encoding
 PT a pesticidal agent such as trypsin modulating coostatic factor compounds
 PT or neuropeptide F compounds, is useful for controlling plant pests -
 XX
 PS Example 2; Page 26; 48pp; English.
 XX
 CC The present sequence is mosquito Trypsin Modulating Coostatic Factor
 CC (TMOF). TMOF is a hormone, which regulates the biosynthesis of the
 CC digestive enzyme trypsin. TMOF is synthesized in the follicular
 CC epithelium of the ovary 2-30 hours after a blood meal and is released
 CC into the haemolymph, binding to the TMOF receptor on the midgut
 CC epithelial cells, signalling the termination of trypsin biosynthesis.
 CC TMOF was used to produce a recombinant plant cell, which is useful for
 CC controlling agricultural pests, in particular insects, and is also useful
 CC for inhibiting the production of digestive enzymes in a pest to control
 CC the pest.
 CC
 SO Sequence 11 AA:
 Query Match 0.7%; Score 9; DB 21; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.74;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 218 PAPPPPPP 226
 Db | | | | | | | | | |
 Db 3 PAPPPPPP 11
 RESULT 11
 ID AAB30628
 ID AAB30628 standard; peptide: 11 AA.
 XX
 AC AAB30628;
 XX
 DT 19-MAR-2001 (first entry)
 XX
 DE Trypsin modulating coostatic factor (TMOF) pesticidal analogue.
 XX
 KW Hormone; trypsin modulating coostatic factor; TMOF; analogue; mosquito;
 KW trypsin; trypsin-like enzyme; metabolic energy; insect pest; pesticide;
 KW digestion; digestive enzyme; flesh fly; flea; sand fly; house fly;
 KW dog fly; coleopteran; lepidopteran; dipteran; blood-sucking insect;
 KW Diptera.
 XX
 OS Synthetic.
 XX
 PN WO200062792-A2.
 XX
 PD 26-OCT-2000.

XX
 PF 04-APR-2000; 2000WO-US08879.
 XX
 PR 21-APR-1999; 99US-0295924.
 XX
 PA (INSE-) INSECT BIOTECHNOLOGY INC.
 PA (UYFL) UNIV FLORIDA.
 XX
 PI Bennett J, Brandt A, Borovsky D;
 XX
 DR WPI; 2001-006951/01.
 XX
 PT New polypeptide is useful for preventing, reducing and eliminating
 PT infestation of area by pests e.g. flesh flies or mosquito larvae,
 PT optionally in combination with e.g. repellent, attractant, acaricide,
 PT fungicide or herbicide -
 XX
 PS Claim 1; Page 58; 87pp; English.
 XX
 CC The present sequence represents an analogue of a peptide hormone named
 CC trypsin modulating coostatic factor (TMOF). Mosquitoes regulate trypsin
 CC and trypsin-like enzymes with this hormone to conserve metabolic
 CC energy. Other insect pests use it for the same purpose. The peptide
 CC functions as a pesticidal compound, which inhibits digestion in pests by
 CC inhibiting synthesis of pests digestive enzymes. The pesticidal
 CC compound is useful for preventing, reducing or eliminating infestation
 CC of geographical areas by an insect population such as flesh flies, fleas,
 CC sand flies, house flies and dog flies. The pesticidal compound is applied
 CC to pest inhabited loci of the geographical area such as the body of
 CC water inhabited by mosquito larvae, or insects such as coleopterans,
 CC lepidopterans, dipterans or blood-sucking insects of order Diptera,
 CC suborder Nematocera, family Colicidae or subfamily Culicinae,
 CC Corethrinae, Ceratopogonidae and Simuliidae.
 CC
 SO Sequence 11 AA:
 Query Match 0.7%; Score 9; DB 22; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.74;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 218 PAPPPPPP 226
 Db | | | | | | | | | |
 Db 3 PAPPPPPP 11
 RESULT 12
 ID AAM39024
 ID AAM39024 standard; peptide: 15 AA.
 XX
 AC AAM39024;
 XX
 DT 27-MAR-1998 (first entry)
 XX
 DE Peptide resembling an SH3 domain binding peptide SEQ ID NO:423.
 XX
 KW Cortactin; SH3 domain; binding peptide; Src homology region 3;
 KW tyrosine kinase; immune response; lymphokine; interleukin 1; Nck;
 KW Abl; P13gamma; p53bp2; Crk; Yes; Grb2.
 XX
 OS Synthetic.
 XX
 PN WO9730074-A1.
 XX
 PD 21-AUG-1997.
 XX
 PF 14-FEB-1997; 97WO-US02298.
 XX
 PR 16-FEB-1996; 96US-0602999.
 XX
 PA (CYTO-) CYTOGEN CORP.
 PA (UYNC-) UNIV NORTH CAROLINA.
 XX

PI Der CJ, Fowlkes DM, Kay BK, Quilliam LA, Rider JE;
PI Sparks AB, Thorn JM;
XX WPI; 1997-424972/39.
DR
XX
XX Src homology region 3 binding peptide - used to activate Src
PT tyrosine kinase(s) and to stimulate immune response by increasing
PT production of certain lymphokine(s), e.g. interleukin-1
XX
XX
PS Claim 22; Page 93; 131pp; English.
XX
XX The present sequence represents a peptide which resembles a Src homology
CC region 3 (SH3) binding peptide. SH3 binding peptides are selected from:
CC (a) peptides which bind the SH3 domain of Crk; (b) peptides which
CC bind the middle SH3 domain of Nck; (c) peptides which bind the SH3
CC domain of Abl; (d) peptides which bind the SH3 domain of Src; (e)
CC peptides which bind the SH3 domain of PLC gamma; (f) peptides which bind
CC the SH3 domain of p53p2; (g) peptides which bind the amino-terminal SH3
CC domain of Crk; (h) peptides which bind the SH3 domain of Yes; and (i)
CC binding peptides can be used in the method to identify inhibitors of
CC their binding to their respective SH3 domains, which could be used to
CC modulate the pharmacological activity of proteins or polypeptide
CC containing the SH3 domain. The peptides can also be used to activate
CC Src or Src-related protein tyrosine kinases, to stimulate the immune
CC response by increasing the production of certain lymphokines, e.g.
CC tumour necrosis factor-alpha and interleukin-1, or to deliver a
CC conjugated molecule to certain cellular compartments containing Src or
CC Src related proteins.
XX
XX Sequence 15 AA;
SQ

Query Match 0.7%; Score 9; DB 18; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.96;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 218 PAPPPEPPP 226
Db 3 PAPPPEPPP 11
XXXXXXXXXXXX

RESULT 13
AAW05474
ID AAW05474 standard; Peptide; 18 AA.
XX
XX
XX AAW05474;
AC
XX
XX 24-FEB-1998 (first entry)
DT
XX
XX SH3-binding peptide bSH3025.
DE
XX
XX Src-homology region 3 domain; human; mouse; SH3 domain; cell growth;
KW cellular signalling element; cellular structural element; malignancy;
KW protein identification; functional domain; protein screening;
KW cellular signal transduction process; binding peptide.
XX
XX Synthetic.
OS
XX
XX MO9631625-A1.
PN
XX
XX 10-OCT-1996.
PD
XX
XX 04-APR-1996; 96WO-US04454.
PF
XX
XX 03-APR-1996; 96US-0630915.
PR
XX
XX 07-APR-1995; 95US-0417872.
XX
XX (CYTO-) CYTOGEN CORP.
PA
XX
XX (UYNC-) UNIV NORTH CAROLINA.
PI
XX
XX Fowlkes DM, Hoffman N, Kay BK, McConnell SJ, Sparks AB;

DR WPI; 1996-465045/46.
XX
XX
XX Identifying polypeptide(s) having specific functional domain (esp.
PT SH3 domain) - comprises detecting selective binding to recognition
PT unit, regardless of sequence homology
XX
XX
XX Example; Fig 13; 174pp; English.
PS
XX
XX AAW05445-W05492 represent Src-homology region 3 (SH2) domain binding
CC peptides. These sequences were used as parts of multivalent recognition
CC unit complexes used in the method of the invention. The method of the
CC invention is for identifying polypeptides containing functional domains
CC of interest (especially SH3 domains). It comprises contacting a
CC multivalent recognition unit (RU) complex with a number of peptides and
CC identifying polypeptides having a selective binding affinity for the RU
CC complex. The method is based on functional similarities and does not
CC rely on sequence similarities. Prior methods only gave limited success
CC for identifying proteins containing an SH3 domain due to the minimal
CC sequence homology among known SH3 proteins. Multivalent RU complexes are
CC particularly suited to screening for polypeptides containing functional
CC domains that are similar to, but not identical in sequence to, the
CC original target functional domain. The new method enables proteins
CC having a common function to be identified. Identification of novel SH3
CC proteins will be useful for a better understanding of cell growth,
CC malignancy, signal transduction processes, etc. New candidate drugs can
CC be identified, and their specificities (e.g. pharmacological activities)
CC can be assessed using the method of the invention.
XX
XX Sequence 18 AA;
SQ

Query Match 0.7%; Score 9; DB 17; Length 18;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 218 PAPPPEPPP 226
Db 5 PAPPPEPPP 13
XXXXXXXXXXXX

RESULT 14
AAB30656
ID AAB30656 standard; peptide; 21 AA.
XX
XX
XX AAB30656;
AC
XX
XX 19-MAR-2001 (first entry)
DT
XX
XX Trypsin modulating oostatic factor (TMOF) pesticidal analogue.
DE
XX
XX Hormone; trypsin modulating oostatic factor; TMOF; analogue; mosquito;
KW trypsin; trypsin-like enzyme; metabolic energy; insect pest; pesticide;
KW digestion; digestive enzyme; flesh fly; flea; sand fly; house fly;
KW dog fly; coleopteran; lepidopteran; dipteran; blood-sucking insect;
KW Diptera.
XX
XX Synthetic.
OS
XX
XX WO200062792-A2.
PN
XX
XX 26-OCT-2000.
PD
XX
XX 04-APR-2000; 2000WO-US08879.
PF
XX
XX 21-APR-1999; 99US-0295924.
PR
XX
XX (INSE-) INSECT BIOTECHNOLOGY INC.
PA
XX
XX (UYFL) UNIV FLORIDA.
PI
XX
XX Bennett J, Brandt A, Borovsky D;
XX
XX WPI; 2001-006951/01.
DR

PT New polypeptide is useful for preventing, reducing and eliminating
 PT infestation of area by pests e.g. flesh flies or mosquito larvae,
 PT optionally in combination with e.g. repellent, attractant, acaricide,
 PT fungicide or herbicide -

XX Example; Page 49; 87pp; English.

CC The present sequence represents an analogue of a peptide hormone named
 CC trypsin modulating oostatic factor (TMOF). Mosquitoes regulate trypsin
 CC and trypsin-like enzymes with this hormone to conserve metabolic
 CC energy. Other insect pests use it for the same purpose. The peptide
 CC functions as a pesticidal compound, which inhibits digestion in pests by
 CC inhibiting synthesis of pests digestive enzymes. The pesticidal
 CC compound is useful for preventing, reducing or eliminating infestation
 CC of geographical areas by an insect population such as flesh flies, fleas,
 CC sand flies, house flies and dog flies. The pesticidal compound is applied
 CC to pest inhabited loci of the geographical area such as the body of
 CC water inhabited by mosquito larvae, or insects such as coleopterans,
 CC lepidopterans, dipterans or blood-sucking insects of order Diptera,
 CC Suborder Nematocera, family Colicidae or subfamily Culicinae,
 CC Corethrinae, Ceratopogonidae and Simuliidae.

SQ Sequence 21 AA;

Query Match 0.7%; Score 9; DB 22; Length 21;
 Best Local Similarity 100.0%; Pred. No. 1.3;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 218 PAPPPIPP 226
 DB 3 PAPPPIPP 11

RESULT 15

AAB30657
 ID AAB30657 standard; peptide: 23 AA.

AC AAB30657;

DT 19-MAR-2001 (first entry)

XX Trypsin modulating oostatic factor (TMOF) pesticidal analogue.

DE Hormone; trypsin modulating oostatic factor; TMOF; analogue; mosquito;
 KW trypsin; trypsin-like enzyme; metabolic energy; insect pest; pesticide;
 KW digestion; digestive enzyme; flesh fly; flea; sand fly; house fly;
 KW dog fly; coleopteran; lepidopteran; dipteran; blood-sucking insect;
 KW Diptera.

OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 2 /note="optionally absent"

FT WO200062792-A2.

PD 26-OCT-2000.

PF 04-APR-2000; 2000WO-US08879.

PR 21-APR-1999; 99US-0295924.

XX (INSE-) INSECT BIOTECHNOLOGY INC.
 PA (UYFL) UNIV FLORIDA.

PI Bennett J, Brandt A, Borovsky D;

DR WPI; 2001-006951/01.

XX New polypeptide is useful for preventing, reducing and eliminating
 PT infestation of area by pests e.g. flesh flies or mosquito larvae,

PT optionally in combination with e.g. repellent, attractant, acaricide,
 PT fungicide or herbicide -

XX Example; Page 49; 87pp; English.

CC The present sequence represents an analogue of a peptide hormone named
 CC trypsin modulating oostatic factor (TMOF). Mosquitoes regulate trypsin
 CC and trypsin-like enzymes with this hormone to conserve metabolic
 CC energy. Other insect pests use it for the same purpose. The peptide
 CC functions as a pesticidal compound, which inhibits digestion in pests by
 CC inhibiting synthesis of pests digestive enzymes. The pesticidal
 CC compound is useful for preventing, reducing or eliminating infestation
 CC of geographical areas by an insect population such as flesh flies, fleas,
 CC sand flies, house flies and dog flies. The pesticidal compound is applied
 CC to pest inhabited loci of the geographical area such as the body of
 CC water inhabited by mosquito larvae, or insects such as coleopterans,
 CC lepidopterans, dipterans or blood-sucking insects of order Diptera,
 CC Suborder Nematocera, family Colicidae or subfamily Culicinae,
 CC Corethrinae, Ceratopogonidae and Simuliidae.

SQ Sequence 23 AA;

Query Match 0.7%; Score 9; DB 22; Length 23;
 Best Local Similarity 100.0%; Pred. No. 1.4;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

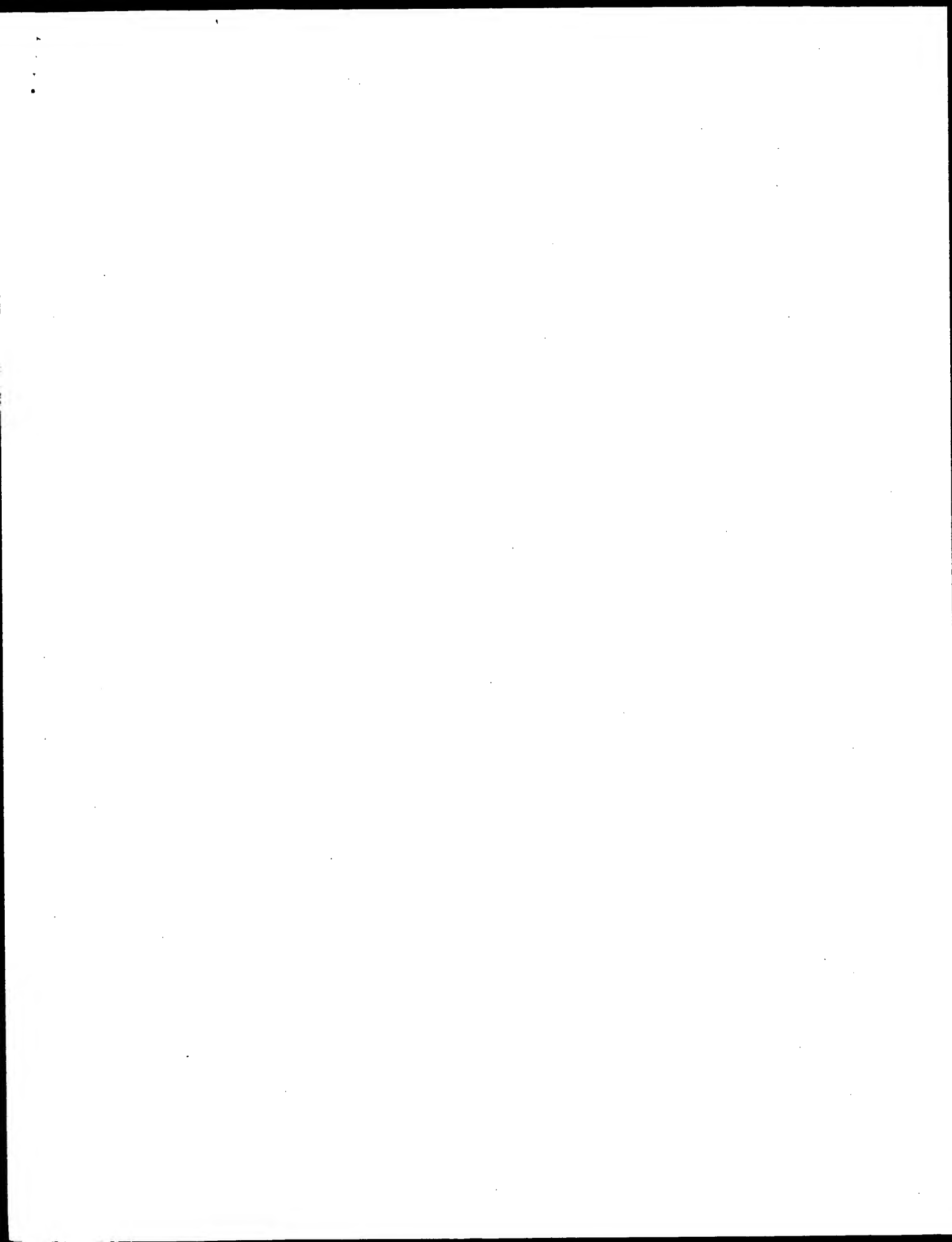
OY 218 PAPPPIPP 226
 DB 4 PAPPPIPP 12

Search completed: July 19, 2001, 16:07:46
 Job time: 206 sec

Fri Jul 20 10:09:29 2001

us-09-501-171a-4.oli.rag

Page 11



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2001, 16:05:20 ; Search time 15.55 Seconds
(without alignments)
1586,957 Million cell updates/sec

Title: US-09-501-171A-4

Sequence: 1 MFARKPPGAAPLGAMPVDDQ.....PYSELNETHSHYSPDSWV 1225

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 193259 seqs, 20144635 residues

Word size : 0

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued Patents, AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCITUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	9	0.7	15	4	US-08-602-999A-423
2	9	0.7	210	1	US-08-234-783-2
3	9	0.7	210	1	US-08-456-907-2
4	9	0.7	210	5	PCT-US95-05523-2
5	9	0.7	429	1	US-08-234-783-4
6	9	0.7	429	1	US-08-456-907-4
7	9	0.7	429	5	PCT-US95-05523-4
8	9	0.7	449	1	US-08-102-942A-4
9	9	0.7	449	1	US-08-102-942A-6
10	9	0.7	539	3	US-08-906-360-1
11	9	0.7	928	1	US-08-204-328-1
12	9	0.7	928	2	US-08-959-638-8
13	9	0.7	928	2	US-08-482-627-5
14	9	0.7	928	3	US-08-801-092-4
15	9	0.7	928	4	US-08-328-673A-8
16	9	0.7	928	5	PCT-US94-10357-2
17	9	0.7	928	5	PCT-US94-10357-3
18	9	0.7	1323	5	PCT-US94-10357-3
19	9	0.7	113	3	US-08-026-138E-4
20	8	0.7	11	3	US-08-105-904B-9
21	8	0.7	11	4	US-08-105-904B-21
22	8	0.7	11	4	US-08-114-877A-9
23	8	0.7	11	4	US-08-114-877A-14
24	8	0.7	15	4	US-08-602-999A-348
25	8	0.7	15	4	US-08-602-999A-352
26	8	0.7	15	4	US-08-602-999A-381
27	8	0.7	18	4	US-08-602-999A-422

28	8	0.7	19	4	US-08-602-999A-408	Sequence 408, App
29	8	0.7	21	3	US-08-630-916A-108	Sequence 108, App
30	8	0.7	25	1	US-08-240-514-56	Sequence 56, App
31	8	0.7	25	2	US-08-612-302A-56	Sequence 56, App
32	8	0.7	223	1	US-08-167-035-43	Sequence 43, App
33	8	0.7	223	1	US-08-208-887A-43	Sequence 43, App
34	8	0.7	223	2	US-08-359-005-43	Sequence 43, App
35	8	0.7	421	4	US-09-093-448-4	Sequence 4, App
36	8	0.7	445	2	US-08-900-148-2	Sequence 2, App
37	8	0.7	472	2	US-08-976-255-17	Sequence 17, App
38	8	0.7	478	2	US-09-040-799-3	Sequence 3, App
39	8	0.7	478	4	US-09-093-448-1	Sequence 1, App
40	8	0.7	478	4	US-09-093-448-2	Sequence 2, App
41	8	0.7	478	4	US-09-093-448-3	Sequence 3, App
42	8	0.7	480	4	US-09-108-020-6	Sequence 6, App
43	8	0.7	549	1	US-08-354-240A-2	Sequence 2, App
44	8	0.7	550	1	US-08-354-240A-4	Sequence 4, App
45	8	0.7	550	1	US-08-354-240A-6	Sequence 6, App

ALIGNMENTS

RESULT 1
US-08-602-999A-423

Sequence 423, Application US/08602999A

Patent No. 6184205

GENERAL INFORMATION:

APPLICANT: SPARKS, Andrew B.

APPLICANT: KAY, Brian K.

APPLICANT: THORN, Judith M.

APPLICANT: OUILIAM, Lawrence A.

APPLICANT: DER, Channing J.

APPLICANT: FOWLES, Dana M.

APPLICANT: RIDER, James E.

TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF

NUMBER OF SEQUENCES: 467

CORRESPONDENCE ADDRESS:

ADDRESS: Pennie & Edmonds

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentln Data: 1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/602,999A

FILING DATE: 16-FEB-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Mistrock, S. Leslie

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 1101-202

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-9741/8864

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 423:

SEQUENCE CHARACTERISTICS:

LENGTH: 15 amino acids

TYPE: amino acid

TOPOLOGY: unknown

MOLECULE TYPE: peptide

US-08-602-999A-423

Query Match 0.7%; Score 9; DB 4; Length 15;

Best Local Similarity 100.0%; Pred. No. 0.43;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 218 PAPPPEPP 226
|||||
Db 3 PAPPPEPP 11

RESULT 2
US-08-234-783-2
; Sequence 2, Application US/08234783
; Patent No. 5622835
; GENERAL INFORMATION:
; APPLICANT: Herlyn, Meenhard
; APPLICANT: Morris, Jennifer
; APPLICANT: Rauscher III, Frank J.
; APPLICANT: Rodeck, Ulrich
; TITLE OF INVENTION: Wt1 Monoclonal Antibodies and Methods of
; TITLE OF INVENTION: Use Therefor
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, PO Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/234,783
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: MST48USA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 210 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-234-783-2

Query Match 0.7%; Score 9; DB 1; Length 210;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 PAPPPEPP 226
|||||
Db 69 PAPPPEPP 77

RESULT 3
US-08-456-907-2
; Sequence 2, Application US/08456907
; Patent No. 5633142
; GENERAL INFORMATION:
; APPLICANT: Herlyn, Meenhard
; APPLICANT: Morris, Jennifer
; APPLICANT: Rauscher III, Frank J.
; APPLICANT: Rodeck, Ulrich
; TITLE OF INVENTION: Wt1 Monoclonal Antibodies and Methods of
; TITLE OF INVENTION: Use Therefor
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Cntr, PO Box 457
CITY: Spring House
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19477

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/456,907
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/234,783
FILING DATE: 28-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: MST48AUSA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 210 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-456-907-2

Query Match 0.7%; Score 9; DB 1; Length 210;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 PAPPPEPP 226
|||||
Db 69 PAPPPEPP 77

RESULT 4
PCT-US95-05523-2
; Sequence 2, Application PC/TUS9505523
; GENERAL INFORMATION:
; APPLICANT: The Wistar Institute of Anatomy and Biology
; TITLE OF INVENTION: Wt1 Monoclonal Antibodies and
; TITLE OF INVENTION: Methods of Use Therefor
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, PO Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/05523
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/234,783
; FILING DATE: 28-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215

REFERENCE/DOCKET NUMBER: MST48PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 210 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-05523-2

Query Match 0.7%; Score 9; DB 5; Length 210;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 218 PAPPPIPP 226
|||||
DB 69 PAPPPIPP 77

RESULT 5
US-08-234-783-4
Sequence 4, Application US/08234783
Patent No. 5622835
GENERAL INFORMATION:
APPLICANT: Herlyn, Meenhard
APPLICANT: Morris, Jennifer
APPLICANT: Rauscher III, Frank J.
APPLICANT: Rodeck, Ulrich
TITLE OF INVENTION: Wt1 Monoclonal Antibodies and Methods of
NUMBER OF SEQUENCES: 7
TITLE OF INVENTION: Use Therefor
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Cntr, PO Box 457
CITY: Spring House
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/234,783
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: MST48USA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 429 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-234-783-4

Query Match 0.7%; Score 9; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 218 PAPPPIPP 226
|||||
DB 58 PAPPPIPP 66

RESULT 6
US-08-456-907-4
Sequence 4, Application US/08456907
Patent No. 5633142
GENERAL INFORMATION:
APPLICANT: Herlyn, Meenhard
APPLICANT: Morris, Jennifer
APPLICANT: Rauscher III, Frank J.
APPLICANT: Rodeck, Ulrich
TITLE OF INVENTION: Wt1 Monoclonal Antibodies and Methods of
NUMBER OF SEQUENCES: 7
TITLE OF INVENTION: Use Therefor
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Cntr, PO Box 457
CITY: Spring House
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/456,907
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/234,783
FILING DATE: 28-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: MST48USA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 429 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-456-907-4

Query Match 0.7%; Score 9; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 218 PAPPPIPP 226
|||||
DB 58 PAPPPIPP 66

RESULT 7
PCT-US95-05523-4
Sequence 4, Application PC/TUS9505523
GENERAL INFORMATION:
APPLICANT: The Wistar Institute of Anatomy and Biology
TITLE OF INVENTION: Wt1 Monoclonal Antibodies and
METHODS OF USE THEREFOR
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Cntr, PO Box 457
CITY: Spring House
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19477

```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/05523
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/234,783
; FILING DATE: 28-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 429 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-05523-4

Query Match      0.7%; Score 9; DB 5; Length 429;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 PAPPPEPP 226
Db 58 PAPPPEPP 66

RESULT 8
US-08-102-942A-4
; Sequence 4, Application US/08102942A
; Patent No. 5726288
; GENERAL INFORMATION:
; APPLICANT: Gall, Katherine M.
; APPLICANT: Glaser, Thomas M.
; APPLICANT: Ito, Caryn Y.
; APPLICANT: Buckler, Alan J.
; APPLICANT: Pelletier, Jerry
; APPLICANT: Haber, Daniel A.
; APPLICANT: Rose, Elise A.
; APPLICANT: Housman, David E.
; APPLICANT: Bruening, Wendy
; APPLICANT: Darveau, Andre
; TITLE OF INVENTION: Localization and Characterization of the
; TITLE OF INVENTION: Wilms' Tumor Gene
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Millitia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/102,942A
; FILING DATE: 02-AUG-1993
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; 
```

```

;
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-5194A2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 449 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-102-942A-4

Query Match      0.7%; Score 9; DB 1; Length 449;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 PAPPPEPP 226
Db 59 PAPPPEPP 67

RESULT 9
US-08-102-942A-6
; Sequence 6, Application US/08102942A
; Patent No. 5726288
; GENERAL INFORMATION:
; APPLICANT: Gall, Katherine M.
; APPLICANT: Glaser, Thomas M.
; APPLICANT: Ito, Caryn Y.
; APPLICANT: Buckler, Alan J.
; APPLICANT: Pelletier, Jerry
; APPLICANT: Haber, Daniel A.
; APPLICANT: Rose, Elise A.
; APPLICANT: Housman, David E.
; APPLICANT: Bruening, Wendy
; APPLICANT: Darveau, Andre
; TITLE OF INVENTION: Localization and Characterization of the
; TITLE OF INVENTION: Wilms' Tumor Gene
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Millitia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/102,942A
; FILING DATE: 02-AUG-1993
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 449 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-102-942A-6

```

Query Match 0.7%; Score 9; DB 1; Length 449;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 PAPPPEPP 226
|||||
DB 58 PAPPPEPP 66

RESULT 10

US-08-906-360-1
; Sequence 1, Application US/08906360
; Patent No. 6013441
; GENERAL INFORMATION:
; APPLICANT: Meruelo, Daniel
; APPLICANT: Pampero, Christine
; TITLE OF INVENTION: MAMMALIAN HUMAN FXI-T1
; FILE REFERENCE: 8105-010
; CURRENT APPLICATION NUMBER: US/08/906,360
; CURRENT FILING DATE: 1997-08-05
; EARLIER APPLICATION NUMBER: 60/023,173
; EARLIER FILING DATE: 1996-08-05
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 539
; TYPE: PRT
; ORGANISM: mammalian
US-08-906-360-1

Query Match 0.7%; Score 9; DB 3; Length 539;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 PAPPPEPP 226
|||||
DB 146 PAPPPEPP 154

RESULT 11

US-08-204-329-1
; Sequence 1, Application US/08204329
; Patent No. 5710255
; GENERAL INFORMATION:
; APPLICANT: SHEPARD, H. M.
; APPLICANT: MEN, SHU F.
; TITLE OF INVENTION: CHARACTERIZATION OF A NOVEL ANTI-P110RB
; TITLE OF INVENTION: MONOCLONAL ANTIBODY
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND & TOWNSEND & CREW LLP
; STREET: TWO EMBARCADERO CENTER, 8TH FLOOR
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: U.S.A.
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/204,329
; FILING DATE: 15-AUG-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/05866
; FILING DATE: 14-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: RENEE A. FITTS
; REGISTRATION NUMBER: 35,136

REFERENCE/DOCKET NUMBER: 16930-000400US

TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 326-2400
TELEFAX: (415) 326-2422

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 928 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-204-329-1

Query Match 0.7%; Score 9; DB 1; Length 928;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 PAPPPEPP 226
|||||
DB 21 PAPPPEPP 29

RESULT 12

US-08-959-638-8
; Sequence 8, Application US/08959638
; Patent No. 5932210
; GENERAL INFORMATION:
; APPLICANT: Gregory, Richard J.
; APPLICANT: Mills, Ken N.
; APPLICANT: Maneval, Daniel C.
; TITLE OF INVENTION: Recombinant Adenoviral Vector and
; TITLE OF INVENTION: Methods of Use
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/959,638
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,673
; FILING DATE: 25-OCT-1994
; APPLICATION NUMBER: US/08/233,777
; FILING DATE: 19-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/142,669
; FILING DATE: 25-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 928 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-959-638-8

Query Match 0.7%; Score 9; DB 2; Length 928;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 PAPPPIPP 226
DB 21 PAPPPIPP 29

RESULT 13

US-08-482-627-5
; Sequence 5, Application US/08482627
; Patent No. 5998134
; GENERAL INFORMATION:
; APPLICANT: Lee, Eva Y-H.P
; APPLICANT: Lee, Eva Y-H.P
; TITLE OF INVENTION: Retinoblastoma Gene - Cancer Suppressor
; TITLE OF INVENTION: and Regulator
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/482,627
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/951,947
; FILING DATE: 28-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; TELEPHONE: (619) 535-9901
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 928 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-482-627-5

Query Match 0.7%; Score 9; DB 2; Length 928;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 PAPPPIPP 226
DB 21 PAPPPIPP 29

RESULT 14

US-08-801-092-4
; Sequence 4, Application US/08801092
; Patent No. 6074850
; GENERAL INFORMATION:
; APPLICANT: Antelman, Douglas
; APPLICANT: Gregory, Richard J.
; APPLICANT: Wils, Kenneth N.
; TITLE OF INVENTION: Tissue Specific Expression of

; TITLE OF INVENTION: Retinoblastoma Protein
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/801,092
; FILING DATE: 14-FEB-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/751,517
; FILING DATE: 15-NOV-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Flits, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 016930-001020
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 703-576-0300
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 928 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
US-08-801-092-4

Query Match 0.7%; Score 9; DB 3; Length 928;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 PAPPPIPP 226
DB 21 PAPPPIPP 29

RESULT 15

US-08-328-673A-8
; Sequence 8, Application US/08328673A
; Patent No. 6210939
; GENERAL INFORMATION:
; APPLICANT: Gregory, Richard J.
; APPLICANT: Wills, Ken N.
; APPLICANT: Maneval, Daniel C.

; TITLE OF INVENTION: Recombinant Adenoviral Vector and
; Methods of Use

; NUMBER OF SEQUENCES: 9

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,673A

```

: FILING DATE: 25-Oct-1994
: CLASSIFICATION: <Unknown>
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/142,669
: FILING DATE: 25-Oct-1993
: APPLICATION NUMBER: US 08/233,669
: FILING DATE: 26-Apr-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Smith, Timothy S.
: REGISTRATION NUMBER: 35,367
: REFERENCE/DOCKET NUMBER: 016930-000920US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 576-0200
: TELEFAX: (415) 576-0300
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 928 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-08-328-673A-8

```

```

Query Match          0.7%; Score 9; DB 4; Length 928;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 218 PAPPPPPP 226
   |||||
Db 21 PAPPPPPP 29

```

Search completed: July 19, 2001, 16:08:05
Job time: 165 sec

